

Psychological Bulletin

CONTENTS

ARTICLES:

- A Reconsideration of the Problem of Introspection... DAVID BAKAN 105
 Rating Scales and Check Lists for the Evaluation
 of Psychopathology..... MAURICE LORR 119
 Recent Studies of Simple Reaction Time.... WARREN H. TRICHNER 128
 Representative vs. Systematic Design in
 Clinical Psychology..... KENNETH R. HAMMOND 150
 Kolmogorov-Smirnov Tests for Psychological
 Research..... LEO A. GOODMAN 160
 Remark on "A Qualification in the Use of
 Analysis of Variance"..... VICTOR H. DENENBERG 169
 Test of Significance for a Series of Statistical
 Tests..... JAMES M. SAKODA,
 BURTON H. COHEN,
 AND GEOFFREY BEALL 172
 Comments on Seeman's Operational Analysis of
 the Freudian Theory of Daydreams..... RICHARD A. BEHAN
 AND FRANCES L. BEHAN 176
 Reply to the Behans..... WILLIAM SEEMAN 178

SPECIAL REVIEW:

- An Evaluation of the Annual Review of Psychology
 (Volumes I-IV)..... LYLE H. LANIER 180

BOOK REVIEWS:

- Shaffer and Lazarus' Fundamental Concepts in
 Clinical Psychology..... GEORGE W. ALNEZ 190
 Klein, Heimann, Isaacs, and Riviere's Developments
 in Psychoanalysis..... ANN MAGARET GARNER 191
 Greene's Measurements of Human Behavior..... EVELYN RASKIN 193
 Baumgarten's La Psychotechnique dans le Monde Moderne..... JOSEF BROSEK 194
 Lawhe's Psychology of Industrial Relations..... JOSEPH W. WHEEL 195
 Powdermaker and Frank's Group Psychotherapy: Studies in
 Methodology of Research and Therapy..... NICHOLAS HOBBS 196
 Gellhorn's Physiological Foundations of Neurology
 and Psychiatry..... WILLIAM A. HUNT 197
 Soodi's Experimental Diagnostics of Drives..... VICTOR C. RABY 198

- BOOKS AND MONOGRAPHS RECEIVED..... 200

Published Bimonthly by the
 American Psychological Association

WAYNE DENNIS, Editor
Brooklyn College

EDWARD GIBSON, Associate Editor (Book Reviews)
Brooklyn College

ROBERT L. THORNDIKE, Associate Editor (Statistics)
Teachers College, Columbia University

LORRAINE BOUTHELET, Managing Editor

The Psychological Bulletin contains evaluative reviews of the literature in various fields of psychology, methodological articles, critical notes, and book reviews. This JOURNAL does not publish reports of original research or original theoretical articles.

Editorial communications and manuscripts should be sent to Wayne Dennis, Department of Psychology, Brooklyn College, Brooklyn 10, New York. Books for review should be sent to Edward Gibson, at the same address.

Preparation of articles for publication. Authors are strongly advised to follow the general directions given in the "Publication Manual of the American Psychological Association" (*Psychological Bulletin*, 1962, 49 (No. 4, Part 2), 389-449). Special attention should be given to the section on the preparation of the references (pp. 432-440), since this is a particular source of difficulty in long reviews of research literature. *All copy must be double spaced, including the references.*

Reprints. Fifty free offprints are given to contributors of articles, notes, and special reviews. Five copies of the JOURNAL are supplied gratis to the authors of book reviews.

Business communications—including subscriptions, orders of back issues, and changes of address—should be sent to the American Psychological Association, 1333 Sixteenth Street N.W., Washington 6, D. C.

Annual subscription: \$6.00 (Foreign \$8.50). Single copies, \$1.50.

PUBLISHED MONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1333 Sixteenth Street N.W., Washington 6, D.C.

Permitted to send class mail matter at the post office at Washington, D.C., under the act of March 3, 1879. Additional entry at the post office at Memphis, Tennessee. Approved for mailing at special rate of postage under the provisions of Act 10-48 Post. (52) provided for in Section 362, act of February 26, 1935, authorized August 6, 1967. Printed in U.S.A.

Copyright, 1968, by The American Psychological Association, Inc.

Psychological Bulletin

A RECONSIDERATION OF THE PROBLEM OF INTROSPECTION

DAVID BAKAN

University of Missouri

It is the purpose of this essay to raise a general question for rethinking in the perspective of modern times. Two related considerations are involved in the motivation to write and publish an essay on introspection as a method for the investigation of psychological phenomena. The first is a sense of society's need for a psychology which is more appropriate to its problems. The second is a conviction that although psychologists should be methodologically careful, they should not afford themselves the luxury of methodological snobbery. There is no investigatory method which is "pure," and which provides an absolute guarantee against the commission of error. If errors be committed, we look to the future for their correctives. In the meantime, and perhaps ultimately, we accept a pragmatic criterion.

It is characteristic in the history of ideas that when some notion is rejected, even for adequate cause, many seemingly associated notions get rejected with it. Often these associated notions may be sound. One of the theses of this essay is that such has been the case with introspection. In the outright rejection of the method of introspection, much that was of considerable value was rejected.

In spite of the avowed rejection of the method, it has stayed with us in several disguised forms. As Boring

has recently indicated, "introspection is still with us, doing its business under various aliases, of which *verbal report* is one" (4, p. 169). Boring seems relatively uncritical of the manner in which we contemporarily avail ourselves of introspection. The argument here is for a careful and avowed use of introspection.

In less disguised form introspection is with us in contemporary clinical psychology. The method of introspection is the method that the patient uses, although there is little avowed recognition of it as the method of the clinician, except perhaps among the psychoanalysts (15). However, "therapy" is coming to be viewed as appropriate training for the aspirant clinician even in non-psychoanalytic contexts.

A HYPOTHESIS CONCERNING THE REJECTION OF THE INTROSPECTIVE METHOD

The rejection of the method of introspection is coincident with the inception of behaviorism in America. The first important behavioristic pronouncement took place in 1913 (20). It is important to understand the immediate antecedents of behaviorism in order to understand the wide popularity that it gained. Boring's comprehensive history makes it unnecessary to recount the involved circumstances associated with the death of classical introspection.

Boring believes that it "went out of style . . . because it had demonstrated no functional use and therefore seemed dull, and also because it was unreliable" (4, p. 174). In the next few paragraphs a hypothesis will be offered to supplement that of Boring.

Psychology was in the throes of the Würzburg-Cornell struggle in the first decade of the twentieth century. The Würzburgers had discovered imageless thoughts; and they themselves hardly knew what to do with them. Titchener, at Cornell, sensed the staggering implications of the Würzburg findings, and struggled desperately to reject them (16).

The psychological literature of the time is in many respects confused, repetitive, and—we might say—anguished. Psychology had, it seemed, got itself into absolutely inextricable difficulties; and there was no one within the introspective movement who had the clarity of vision to go beyond these difficulties. Watson, for all the limitations that we may ascribe to him, had clarity and offered a program which psychologists could follow.

Let us briefly examine the nature of some of the Würzburg findings. They discovered that thought was possible without images; and that thought was guided by states variously designated by the terms *Aufgabe*, *Bewusstseinslage*, and *determinierende Tendenz*. The favored method was the *Ausfragemethode*. Mayer and Orth (11) used the method of free association to a verbal stimulus, instructing the subject to report *everything* that went on between the hearing of the stimulus word and the making of the response. Messer (14) finds himself forced to posit *unconscious* processes underlying the processes of thought. Ach (1, 9)¹ in-

troduces the concept of the will, i.e., motivation, as guiding the thought processes; he uses a probing investigatory procedure; and he uses hypnosis. Bühler (6) indicates that it is important, in the study of the thought processes, to empathize and sympathize with the subjects engaged in this kind of experimentation.

Then, the problem is dropped like the proverbial hot potato. Külpe, the leading figure in the Würzburg movement, leaves Würzburg and goes to Bonn in 1909, and the work practically ceases. Bühler posthumously publishes Külpe's lectures which, according to Boring, "contain a pretty complete system of psychology. But the chapter on thought was missing! Bühler said that Külpe had not been lecturing on the topic" (3, p. 407).

In the light of the foregoing, and in the light of what we have learned from psychoanalysis, a rather simple explanation suggests itself. These investigators were using themselves and each other as subjects. They had struck the unconscious, and particularly unconscious motivation, and had to probe it if they were to make any headway. However, as we know today, probing the unconscious tends to generate anxiety and resistance; and these investigators simply were not prepared to undergo the necessary personal trials involved. Boring (4, p. 186) suggests a relationship between the Würzburg school and Freud, but makes little of it.

Psychology had two possible alternatives: either to widen its investigations to take account of and to study the role of unconscious motivation on the thought processes, or to detour. Academic psychology detoured; and detoured in two ways:

¹ The writer could not locate a copy of Ach's book. This sentence is based on Humphrey's (9) summary of Ach's work.

It detoured by way of behaviorism, completely rejecting (at least avowedly) the whole method of introspection, and it detoured by way of gestalt psychology. The former dropped the whole concept of mind, conscious and unconscious. The latter adopted as a basic principle that whatever introspection is done should be *naïve* introspection, with no probing and no analysis, thus presenting intrusion upon the unconscious.

A BASIC DISTINCTION FOR INTROSPECTION

Perhaps one of the most important distinctions necessary for the understanding of the nature of introspection is the classical one between the experience and that of which the experience is. It is the distinction which is contained in the classical one of *Kundgabe* versus *Beschreibung* (4). It is the distinction which is indicated by the concept of stimulus-error (2). It is the distinction which the psychoanalyst makes when he concerns himself primarily with a memory, as contrasted with the event to which the memory presumably refers.

The distinction is somewhat difficult to grasp when we deal with perception. Let us consider a simple experience reported as "I see a book." From the point of view of this distinction it is one or another of two reports: "I *see* a book," or "I see a *book*." In the first instance it is a report of experience as experience. In the second instance the reference is to the object rather than to the experience of the object. The first can be true, and the second false, as, for example, in an hallucination.

The distinction is easier to make when we consider something like anxiety. It is hard to make when the experience involves an external stim-

ulus. It is of interest that when Washburn made her presidential address before the American Psychological Association (19) in 1921 she felt that it was necessary to say that introspection is proper only where there is an external stimulus. This, she believed, would endow introspection with "objectivity"—an unfortunate semantic identification of "object" with "objectivity." It is here, probably, when the Watsonian noose was drawing very tight around the neck of introspection, that introspection surrendered the very thing which was its major merit. Introspection has its maximum value on those very experiences for which there may be no conspicuous physical stimuli, such as grief, joy, anxiety, anger, depression, exhilaration, etc.

THE PROBLEM OF LANGUAGE AND COMMUNICATION

A major criticism which has been leveled against the method of introspection is that the *data* of introspection are not public. In the case of overt behavior it is possible, at least in principle, for two observers to observe a given phenomenon simultaneously. This has sometimes been referred to as the criterion of publicity; and it has been said that data are not acceptable unless this criterion has been satisfied (again, at least in principle).

That introspective data are not public in this sense is not to be questioned. What is to be questioned is whether the criterion is essential. What is the value of the criterion of publicity? Its value, presumably, inheres in the conviction that it avoids error and provides for verification. However, can we not have verification without publicity? Let us consider one of the most acceptable kinds of investigatory procedure from

this point of view, the conditioning experiment. There is no way of verifying Pavlov's experiments today by having another observer watching them, since, to say the least, Pavlov's dogs are probably all quite dead. In order to verify Pavlov's findings we would have to get other dogs. Furthermore, the fact that two people could have stood by to count the number of drops of saliva is quite irrelevant. If the criterion of publicity is not met by introspection, it is not really very serious as long as each scientist has, so to speak, at least one "dog" whom he can observe directly.

The crisis which was generated by disparate results from Würzburg and Cornell, with the one finding imageless thoughts and the other not finding them, was hardly adequate reason for the total rejection of introspection. Disparate results from different laboratories are usually provocative of further investigation, rather than the occasion for dropping the problems, the methods, and the fundamental points of view involved. The failure of the introspective method to satisfy this naive criterion of publicity could hardly have been the real reason for the rejection of introspection as a method.

A more important problem is the possibility of publicity, not of the data, but of the report. Even though the process of introspective observation is, in a sense, private, the information gleaned from the observations must be public. This raises the question of language and communication. There are two questions that may be asked in connection with language with respect to introspection: First, if we relate our introspections to one another, would we understand one another? Second, if we do understand one another, how does this come to pass? If the answer to the first question is to any degree affirmative, then

to that extent is the criterion of publicity of report satisfied.

For the answer to the first question we appeal, at the very least, to common sense. If we hear a person say, "I am sorry," or "I am worried," or "I feel sick," etc., there is hardly any question but that we understand what he means. There are times when we may not *believe* him; but the possibility of fraud, intentional or unintentional, or of lack of precision exists with respect to any methodology. The fact is, however, that we understand him.

The answer to the second question now becomes a matter for empirical investigation. This is not the place to enter into a detailed discussion of the psychology of language learning. However, it is extremely pertinent to indicate that the theory of language learning implicit in contemporary behavioristics is much more simple than is consistent with the facts. This implicit theory may be roughly characterized as follows:

The teacher holds up a ball and says, "Ball." The learner repeats, "Ball." The learner then, presumably, comes to "know" the meaning of the word. Certainly the theory is stretched to the breaking point when presented with the fact that we all fairly well understand the meanings of words such as "sorrow," "feeling," "nausea," "if," "but," etc.

INTROSPECTION AS RETROSPECTION

In 1921 Titchener (18) wrote an essay which, in part, attempted to present to English-speaking readers some of the contributions of Franz Brentano. In the judgment of the writer, Brentano is one of the most important figures in the history of psychology. The major work of Brentano with respect to psychology (5), has not, as far as could be determined by the writer, been translated

into English. Of Brentano and Wundt, Titchener wrote: "The student of psychology, though his personal indebtedness be also twofold, must still make his choice for one or the other. There is no middle way between Brentano and Wundt" (18, p. 108). For the most part, the choice of the classical introspectionists was for Wundt. Wundt and Brentano published their major psychological works at about the same time. Two major schools of thought issue from Brentano. One is the already-mentioned Würzburg school. The other is psychoanalysis, with Brentano having been the only academic psychologist under whom Freud studied (12, 13). Psychoanalysis, however, differed from the Würzburgers with respect to a readiness to face the unconscious. It may have been easier for Freud to break through to the unconscious because it was not his own unconscious but the unconscious of his patients. It was only secondarily that Freud used himself as subject. The Würzburgers, on the other hand, used themselves and each other as subjects.

Brentano, Külpe, and Freud conceived of introspection not as of the present, but the past. They took seriously what was then a common observation that introspection at the moment an experience is taking place changes the character of that experience. If we are interested, say, in anger, then introspection at the moment of anger tends to reduce the anger. It is only when anger is past that it can be properly examined. Using the method of introspection, thus avowedly retrospectively, makes it possible to examine psychological phenomena which cannot readily be elicited in the laboratory, except perhaps with very great ingenuity.

This difficulty of the introspection of Wundt and Titchener was ade-

quately recognized by McDougall. He wrote: "Experimental introspection has obvious limitations. Many of our most vital and interesting experiences, such as grief or joy or fear or moral struggle, cannot be induced at will, except perhaps, in very slight degrees. And, under the most favourable conditions, introspection of our more vivid and vital experiences is difficult, because we are apt to be primarily interested in the events of the outer world in which we are taking part, if only as observers. Then again the very act of introspection does to some extent modify the experiences we wish to observe and describe; so that in introspecting we partially defeat our own purposes" (10, p. 4).

Thus, the type of introspection which was advocated by Titchener, and which was the object of attack by the anti-introspectionists, was a type which, by its nature, could not attack the important aspects and kinds of experience. The cry that a psychology was wanted which would have some usefulness was completely justified when the object of attack was the kind of introspection advocated by Titchener.

ERRORS OF INTROSPECTION

A characteristic of good science is that it is ever alert to the possibility of the commission of systematic types of errors. One of the major criticisms which has been leveled against introspection is that its results are untrustworthy. In the following few paragraphs a brief attempt will be made to examine the problem of the trustworthiness or validity of introspective reports.

There is a respect in which introspective observations are more trustworthy than observations made by the use of the sense organs. Sense organs may be defective. Sense or-

gans are subject to illusion. Observations made with the sense organs are subject to the accidents of angle of regard, kind of illumination, noise level, etc. In the last analysis, the sense organs are subject to hallucination. Introspection is a method which does not involve the sense organs in the usual fashion, and therefore all of the error tendencies associated with the sense organs do not exist for introspection.

However, introspection has associated with it other sources of error. But even at this date, we have achieved a certain amount of progress in isolating them. We know about the stimulus-error. We are aware of the tendency to suppress data (repression), of the tendency to supply socially acceptable data in place of other data (distortion, rationalization, displacement, etc.). But, insofar as we are aware of these error tendencies we can take precautions against their commission. In this respect introspection is no different from any other set of methods in science. To be aware, for example, of the tendency toward rationalization stimulates us to challenge our introspective findings to determine whether they have resulted from the rationalization process. It is a matter of time and careful work to discover other error sources. We have discovered suggestion, cultural determination, ethnocentrism, etc.; and the list will probably lengthen as our experience with the method is enlarged.

A FUNDAMENTAL DIFFERENCE BETWEEN CLASSICAL INTROSPECTION AND PSYCHOANALYSIS

Psychoanalysis has one major limitation with respect to our purposes which was not present in classical

introspection. This is that the major objective of psychoanalysis is therapy.² The major objective of the classical introspectionists was the acquisition of knowledge. This is a fundamental difference.

Essentially, what is being advocated in this paper is the use of the psychoanalytic *method* with the *objective* of the classical introspectionists.

It has been indicated that what is being advocated in this paper is partly on the grounds of the need for a science of psychology which has practical implications. However, there is an old lesson in the history of science of which we avail ourselves. Whereas knowledge may have practicality as its ultimate objective, it has been found that we sometimes do better, both practically and theoretically, if we temporarily forsake the practical objective.

In taking the objective from the classical introspectionists, it is necessary to make some modification in the psychoanalytic procedure. Although the investigator should be "free" in his associations, he should not permit himself to wander too far from the subject under investigation. His associations should stay under the influence of the task at hand. Of course, as in any investigation, decisions of relevance have to be made, and sometimes only a dim intuition dictates the nature of these decisions. Although there is no a priori method for determining relevance, the investigator should always attempt to keep in mind that he is serving science primarily, and himself secondarily.

² This is true even though Freud did envisage that "the future will probably attribute far greater importance to psychoanalysis as the science of the unconscious than as a therapeutic procedure" (7, p. 673).

A "MINIATURE" INVESTIGATION OF
THE RETENTION AND REVELATION OF
SECRETS BY THE METHOD OF RETRO-
SPECTIVE ANALYSIS

In accordance with what has already been said the writer attempted to conduct an investigation of the kind suggested. It is a "miniature" investigation in that it was conducted only over a very short period of time. It was conducted for five days for about an hour and a half each day.

There were several reasons for the choice of the topic, retention and revelation of secrets. One of these is that the topic seemed to be one which is more amenable to introspective investigation than to other methods. By its very nature a secret is something which may not reflect itself in overt behavior. The latter almost constitutes a definition of a secret. Another reason for the choice of the topic is that it seems to be a fundamental one for any kind of introspective investigation. It seemed important to obtain information concerning the nature of secret retention and secret revelation before very much progress could be made with other topics. A third reason was that the topic seemed to lie close to the oft-stated objective of psychology as being prediction and control of human behavior.

The procedure simply involved sitting down to the typewriter and typing whatever came, after the decision concerning the topic was made. The choice of the typewriter was made primarily on the basis that the writer has found himself to be more fluent this way than either writing by hand or talking into a recording machine.

By virtue of the nature of the subject chosen, the writer attempted to write "as though" the material would

never be released. Under any circumstances, even if this was a myth, the sense of the possibility of editing was not mythical. At the moment the writer does not consider it wise to release the protocol. However, one example will be given. The following is taken from the record with some editing:

... What is one of the secrets such as *thee and me* have? I once talked to a professor of zoology at lunch about the academic life. He commented that over the head of every academician hangs a sword on a thin string. No matter how much you do, you never feel that you are doing enough. I am reminded of Freud's dream of Irma's injection. He says, "I am always careful, of course, to see that the syringe is perfectly clean. *For I am conscientious.*" The italics are mine. If he felt that he were really conscientious, if he had no feelings of shortcomings in this connection, why did he have to protest that he *was* conscientious? The guilt of lack of conscientiousness haunts most of my friends. My lack of conscientiousness is my "secret." But here I find myself confessing to lack of conscientiousness. But I was not able to do so until I was able to remember something which would make it possible for me not to have my guilt alone. I brought up the zoology professor. When I wrote the above line about him I hesitated for a moment on the question of whether or not to use quotation marks, or to write it in the way that I did. The quotation marks would have had to come, in all honesty, after the word "string." I wrote on, however, "No matter how much you do, you never feel that you are doing enough." This is what I would have liked him to have said. I added it to give the impression that he had said it, but not quite lying about it.

I think that what has been said above can be generalized. *We are more prone to confess a secret guilt when we can believe that others have the same secret guilt . . .*

The general pattern involved in this kind of writing is that of an oscillation between a free expressive mood and an analytic mood, with the free expression being the subject of the analysis. The question of what a given item of free expression might mean with respect to the major topic

under investigation was repeatedly asked.

In the course of this investigation a series of propositions, including the italicized one above, were formulated. This list can be considered to be the *yield* of this "miniature" investigation:

1. (Given above.)
2. Persons with a secret guilt tend to create situations in which they can "see" that others have the same secret guilt.
3. A secret is a secret by virtue of the anticipation of negative reactions from other people.
4. A secret is maintained in order to maintain some given perception of one's self in others.
5. Persons who associate with one another in the context of a larger group, who have a secret from that larger group, will create a metaphorical or otherwise cryptographic language in which to discuss the secret.
6. In order to conceal a secret one may tend to reveal a fabricated "secret," or a less-secret secret, in order to generate the impression that one is being open and frank.
7. One of the important secret areas in our culture is in connection with our intellectual limitations.
8. When an individual has a secret he will attempt to "protest" that the opposite is the case, if the secret has an opposite.
9. The revelation of a secret may involve the attempt to generate the impression that one is telling a joke, to achieve the double purpose of revelation on the one hand and disbelief on the other.
10. In the revelation of a secret one may attempt to generate the impression that one degrades one's self in one's own eyes, in order to reduce the degradation that one anticipates will be the reaction of others to the revelation.

11. If A knows a secret about B, and B knows a secret about A, and if A discovers that B has revealed A's secret, then A will be inclined to reveal B's secret.

12. If an individual changes his group identification from Group A to Group B, and if Group A has a secret which it keeps from Group B, that individual will be inclined to reveal Group A's secret to the members of Group B.

DISCUSSION OF THE "MINIATURE" INVESTIGATION

The simple fecundity of the method soon became evident. After the decision was made to attempt it and a brief beginning was made, it became apparent that this was, to use a term from the vernacular, a veritable mine of information. Essentially it capitalizes on the fact that the investigator has had twenty or thirty or forty or fifty or sixty or seventy years for the collection of various kinds of information. Certainly one of the defects of this kind of data collection is that it is not systematic in the usual way in which we understand this term. Yet it is the result of years of trial and error, of a kind which most laboratory types of investigation do not generally get. It may be argued that these data have been uncritically gathered. This is a valid point. However, the necessary criticality can be supplied in the course of the investigation itself.

This kind of investigation can be severely hampered by what may be loosely designated as "ethical" considerations. Let us consider, for example, the third proposition enumerated above, that a secret is maintained in order to maintain a given perception of one's self in others. By virtue of the intimate connection between ethics, in this larger sense, and the kind of data which may become

the subject of an introspective investigation, it is extremely important that the investigator attempt, to the degree that he can, to divest the investigation of ethical considerations. Methodologically this divestiture may involve a preliminary investigation of the ethical considerations themselves. Also, it must be added that, for some kinds of problems to be investigated by these methods, less may be required in the way of preliminary investigation than for other problems. However, for the investigation of any problem by these methods, a scientific and objective attitude is a prerequisite.

One of the major merits of this kind of an approach is that it studies the phenomena of psychology *directly*, in a manner which is rarely the case in most psychological investigations. Actually, the kind of material which issues from an introspective investigation of the kind being advocated is *presupposed* in many other psychological investigations. Consider for the moment the "lie" scale of the Minnesota Multiphasic Personality Inventory. The test presumably "gets at" the kind of thing which has been investigated in the investigation on secrets cited above. However, the items of this scale were selected because they would presumably be answered negatively by persons who were trying to put themselves "in the most acceptable light socially" (8). This *presumes*, with little qualification, the content of the fourth proposition, as well as about 15 preconceptions concerning the meaning of social acceptability. (There are 15 items in the "lie" scale.)

Furthermore, had the makers of the MMPI critically examined the nature of secrets in the way in which it has been begun in the above investigation, they would have seen that there are other dynamics of lying, in addi-

tion to the one of which they did avail themselves. For example, proposition 6 indicates that a certain amount of truth-telling may simply be a device for "covering up" one or more other lies. It may well be that the operation of the dynamic indicated by proposition 6 acts to depress the "lie" score when lying is really taking place. A full awareness of the kind of thing that issues from such an investigation can greatly enhance the effectiveness even of pencil-and-paper tests.

From a more theoretical point of view, if we seriously accept the mission of psychology as being that of the prediction and control of human behavior, the psychology of secrets is an important link in the chain of psychological findings and theory. Investigators, no matter what they are investigating, must be cognizant, at the very least, of the possibility of dissemblance when they use human subjects. To predict and control an individual's behavior it is important to know, for example, his group-identifications, his objectives, his values, etc. Many of these items of information are secret. They may even be secret to the subject himself. And under any circumstances they are not items of information which will be revealed readily. Thus, until psychologists develop a rather full understanding of the dynamics of this phenomenon, the ignorance of this phenomenon will stand in the way of other investigations.

What has been said in the above paragraph would be considerably less cogent if the phenomenon of the secret played only a small role in connection with other phenomena. However, secrets play their most important role in those phenomena which are most vital. A psychology that seeks to understand these vital phenomena must have an appreciation of the

phenomena of secret retention and secret revelation. Whether we are interested in the problems of marriage, industrial management, leadership, prejudice, loyalty, delinquency, international affairs, politics, military strategy, litigation, business practices, economics, etc., the psychology of secrets is extremely pertinent. And, the psychology of secrets yields most effectively to the method which is being proposed.

It might be indicated that an adequate psychology of secrets would represent an extremely important contribution to our country in its present state. An adequate psychology of secrets would provide us with an insuperable advantage over our enemies and potential enemies. Concretely, for example, we could use the information about the psychology of secrets for enhancing the skill of the military interrogator in extracting information from prisoners of war; and we could teach our own soldiers about possible techniques that may be used against them in the attempt to extract information from them, so that they may be prepared for them. Also, an adequate psychology of secrets might make quite unnecessary some contemporary political investigatory practices which enjoy some popularity only in lieu of more scientific devices.

Although the psychology of secrets is perhaps a central and basic one associated with the method, investigations could and should be pursued with great profit on other problems. Thus, for example, problem solving and decision making can and should be investigated by the method of retrospective analysis. Investigations on status, power, anxiety, fear, aggression, aesthetic experience, learning, communication, memory, concept formation, perception, judgment,

charity, loneliness, betrayal, etc. could and should be carried out to enhance our understanding of these phenomena.

THE PROBLEM OF VALIDITY OF THE FINDINGS FROM THE "MINIATURE" INVESTIGATION

Perhaps the critical question in the mind of the reader up to this point is that of the "validity" of the findings of an introspective investigation. The problem of validity has already been discussed previously, but somewhat abstractly. In this section, the problem will be dealt with somewhat more concretely, with the findings of the "miniature" investigation before us.

The propositions which issued from the "miniature" investigation are, at least, what may be considered to be "hypotheses" for investigation by other methods. Thus, at the very least, the method may be recommended as a device for systematically getting hypotheses as contrasted with, say, the casual reaching out for a pair of variables and hypothesizing a relationship between them.

Again, as has already been indicated, it may be used as a method whereby an investigator can bring his presuppositions concerning an investigation to *formulation*; where he can critically examine his presuppositions; and where he might be helped in conceiving of other presuppositions against which he can contrast the ones he is using. Or, the method could be used as a device whereby an investigator, having got some experimental results which he cannot understand, provokes his imagination to arrive at some kind of an explanation of his results. The *deliberate* and *avowed* adoption of the method would be extremely helpful in these respects.

However, the writer believes that the method warrants more than this. As has been indicated, the method has a *directness* which is not to be found in any other method of investigation of psychological phenomena. In any investigation each thing which lies between the phenomenon and the data is a source of error. These sources of error are minimized by the method which is being proposed. All errors such as failure of the subject to cooperate (e.g., rehearsal when instructed not to do so in studies on reminiscence), dissemblance, failure to comprehend instructions, refusal to believe the expressions of the investigator's avowed intentions, fear of hidden—or manifest—microphones, lack of skill on the part of the subject (e.g., fixating a point in a vision experiment), refusal to take a "naive" attitude (in gestalt experiments), the lack of control over human subjects (e.g., subjects in problem-solving studies already knowing the solutions to problems but not informing the investigator), subjects knowing the intention of the investigator (e.g., subjects knowing that the experimenter is interested in demonstrating a relation between frustration and aggression, and therefore concealing their felt aggression), etc. are minimized in this kind of investigation.

The propositions which were yielded by the "miniature" investigation also have a certain kind of self-evidence associated with them. They elicit the "of course" response. Some of the propositions may require further specification and further qualification. Nevertheless, they are in some sense obvious. It is the sense of self-evidence which is associated, perhaps, with the axioms of Euclidean geometry. The nature of self-evidence is, of course, an extremely

difficult problem and perhaps more properly falls in the province of the philosopher. Or, perhaps, self-evidence is a problem to be investigated by the very methods which are here proposed. However, whatever the ultimate nature of self-evidence may be, there is a sense in which the results of an introspective investigation are of this type.

Now, of course, the matter of self-evidence may be challenged by the question: Self-evident to whom? In one respect this is a valid question. But in another respect it is not. It is valid in that if we are to know that it is self-evident it must be self-evident to someone. However, when the mathematician uses the term *self-evident* he means something which is intrinsic to the proposition, rather than something dependent upon the reader or the hearer of the proposition. For the mathematician it is the self-evidence of the proposition which makes it possible for the person to see the self-evidence, rather than the reverse. It is this characteristic which is shared by introspective propositions.

As a matter of fact, some of the propositions which issued from the "miniature" investigation seem to partake of greater self-evidence than others. Thus, for example, proposition 4 seems to be quite self-evident, whereas proposition 5 seems to be somewhat less self-evident. And even the seeming self-evidence of proposition 4 may be quite culture bound. However, what has been reported is only an extremely limited investigation, only a beginning and only a sample. Nevertheless, what has been presented is enough to suggest the possibility of achieving the kind of self-evidence that has been indicated.

Two related, but distinguishable, problems are those of replication and

generality. Can such an investigation be replicated? The answer is affirmative, although the difficulties of replication should be recognized and account should be taken of them. If an investigator attempts to replicate his own investigation at another time, he will inevitably be under the influence of what he has already done. In replicating such an investigation, the very replication itself should come under the scrutiny of the investigator. He should challenge, for example, his personal identification with the results he has already obtained, and prepare himself for finding both novelty and contradiction with respect to his earlier investigation. If one investigator is interested in replicating the investigation of another investigator, he should carefully take into account the possibility of suggestion, or of his willingness to accept the results of the earlier investigator (particularly if the first investigator has prestige for the second investigator). He should take careful cognizance of possible motivation for showing the earlier investigator to be in error, etc. In some instances it may be extremely worth while to investigate some topic without reading the results of the earlier investigation until the completion of the second investigation, and the making of a comparison later on. Carefully controlled experimentation to determine possible effects of suggestion, for example, is extremely feasible.

The generality of the results of such an investigation is a somewhat more difficult problem, but it is a difficulty which is not unique to introspective investigation. One investigator's results can be compared with another investigator's results, so that the problem of uniqueness with respect to a single investigator is vitiated. However, one may ask, in the

event of consistency of results among a group of investigators, may the findings not be unique to a group of persons all of whom are introspective investigators? There is no easy answer to this problem. However, we face the same problem in other investigations. May not the results of studies in rote learning be largely unique to college sophomores? May not the results of studies in, say, secondary reinforcement be unique to rats, or more particularly laboratory rats, or even more particularly white laboratory rats, or still more particularly tamed white laboratory rats, etc.? May not all findings concerning mental abnormality be unique to mentally abnormal persons *contacted* by investigators, and may not these very contacts be a major determinant of the findings?

The answer, of course, to each of these questions is contingent upon some decision concerning relevance, a decision that has to be made in connection with any investigation. Actually, the kind of investigation being advocated has an advantage in this respect over other kinds of investigation. For, in an introspective investigation the very decisions concerning relevance can come under the same scrutiny as the phenomena being investigated.

POSSIBILITY AS A FINDING

The argument concerning the validity of the findings from an introspective investigation thus far has been concerned with validity in the usual sense, i.e., the argument has been concerned with the truth or falsity of propositions which issue from an introspective investigation.

There is, however, a value to such propositions which is over and beyond their validity. This is their *possibility* rather than their truth or

falsity. The knowledge that a certain dynamic is *possible* enhances the sensitivity of the psychological observer. To make this point concrete, let us again consider the military interrogation situation. Suppose that the interrogator is interested in determining the contents of some supplies which have been moved in by the enemy. Now suppose that the prisoner being interrogated knows these contents but does not wish to reveal them. The prisoner *may* avail himself of the dynamic indicated by proposition 6 (the revelation of less secret secrets in order to generate the impression that he is being open and frank), and inform the interrogator at length about a great number of lesser secrets, but not the nature of the supplies. He may say, "I will tell you everything that I know, but I do not know what was in those trucks." An interrogator who was not aware of the *possibility* of proposition 6 might be lulled into believing the man. The interrogator might say to himself, "he is evidently telling all that he knows." On the other hand, an interrogator who is aware of the possibility of the action of the dynamic indicated by proposition 6 would be aware of the possibility of this kind of deception, and would be less likely to be "taken in."

Insofar as at least one person can contrive such a device for deception, then such a device is *possible*, and some other individual *may* have conceived of it and *may* be making use of it. The truth of the proposition in this respect becomes quite secondary. What is important, simply, is that someone thought of it; and if one person thought of it, other persons might think of it.

In this respect psychologists can make a major contribution to society not only by rendering to society

established *truths*, but also by rendering to society established *possibles* with respect to psychological dynamics. In the matter of prediction and control of human behavior, a knowledge of what an individual might *possibly* do, or *possibly* feel, or *possibly* think places us well on the way toward the achievement of our objective. Given a detailed knowledge concerning the possibles we can act in such a fashion as to discourage some from becoming actualities, and to encourage others into becoming actualities. The pragmatic usefulness of knowledge of possibles extends from the clinical situation to world affairs.

As has been suggested, these possibles may indeed turn out to be truths in the larger and more scientific sense. But even if they fail to meet the criteria for general scientific propositions, they have value in the sense indicated above.

SUMMARY

It is the purpose of this essay to raise the question of the appropriateness of the method of introspection for rethinking in the perspective of modern times. Some features of the history of introspection in the first decade of the twentieth century have been pointed to. The hypothesis is advanced that introspection was dropped because the classical introspectionists had come to a point where they would have had to probe the unconscious to make any progress. From psychoanalysis we have learned that probing the unconscious generates anxiety and resistance. The old distinction between an experience, and that of which the experience is, is again made. The problems of communication of introspective observations are discussed. It is claimed that introspection should

be retrospective, consistent with the position of Brentano, Külpe, and Freud. The problem of errors in introspection is discussed. The position is advanced that retrospective analysis should take its *objective* from the classical introspectionists and its *method* from the psychoanalysts, with some modification. A "miniature" investigation on the psychology of secret retention and secret revelation is described. The implications of the investigation both as an investigation in its own right and as an example of the method are discussed. The prob-

lem of the validity of the findings is discussed. The problems of replication and generality are discussed. The propositions which issue from such an investigation can be viewed as having three values: first, as hypotheses to be investigated by other methods, and otherwise to supplement other methods; second, as propositions which have a certain kind of self-evidence associated with them; third, as *possibles* which can give us real assistance in achieving the objectives of prediction and control of human behavior.

REFERENCES

1. ACH, N. *Über die Willenstätigkeit und das Denken*. Göttingen: Vandenhoeck und Ruprecht, 1905.
2. BORING, E. G. The stimulus-error. *Amer. J. Psychol.*, 1921, **32**, 449-471.
3. BORING, E. G. *A history of experimental psychology*. (2nd Ed.) New York: Appleton-Century-Crofts, 1950.
4. BORING, E. G. A history of introspection. *Psychol. Bull.*, 1953, **50**, 169-189.
5. BRENTANO, F. *Psychologie vom empirischen Standpunkte*. Leipzig: Duncker und Humblot, 1874.
6. BÜHLER, K. Tatsachen und Probleme zu einer Psychologie der Denkvorgänge: I. Ueber Gedanken. *Arch. ges. Psychol.*, 1907, **9**, 297-365.
7. FREUD, S. Psychoanalysis: Freudian School. *Encyclopedia Britannica*, 14th Ed., v. 18, pp. 672-674.
8. HATHAWAY, S. R., & MCKINLEY, J. *Minnesota Multiphasic Personality Inventory manual*. New York: Psychological Corp., 1951.
9. HUMPHREY, G. *Thinking*. New York: Wiley, 1951.
10. McDUGALL, W. Prolegomena to psychology. *Psychol. Rev.*, 1922, **29**, 1-43.
11. MAYER, A., & ORTH, J. Zur qualitativen Untersuchung der Association. *Z. Psychol. Physiol. Sinnesorg.*, 1901, **26**, 1-13.
12. MERLAN, P. Brentano and Freud. *J. Hist. Ideas.*, 1945, **6**, 375-377.
13. MERLAN, P. Brentano and Freud—a sequel. *J. Hist. Ideas.*, 1949, **10**, 451.
14. MESSER, A. Experimentell-psychologische Untersuchung über das Denken. *Arch. ges. Psychol.*, 1906, **8**, 1-224.
15. REIK, T. *Listening with the third ear*. New York: Farrar & Strauss, 1948.
16. TITCHENER, E. B. *Lectures on the experimental psychology of the thought-processes*. New York: Macmillan, 1909.
17. TITCHENER, E. B. Description vs. statement of meaning. *Amer. J. Psychol.*, 1912, **23**, 165-182.
18. TITCHENER, E. B. Brentano and Wundt: empirical and experimental psychology. *Amer. J. Psychol.*, 1921, **32**, 108-120.
19. WASHBURN, MARGARET F. Introspection as an objective method. *Psychol. Rev.*, 1922, **29**, 89-112.
20. WATSON, J. B. Psychology as the behaviorist views it. *Psychol. Rev.*, 1913, **20**, 158-177.

Received May 18, 1953.

RATING SCALES AND CHECK LISTS FOR THE EVALUATION OF PSYCHOPATHOLOGY

MAURICE LORR

Veterans Administration, Washington, D. C.

The use of check lists, charts, and rating scales for the objective recording and later evaluation of change in the behavior and symptoms of psychiatric patients is not new. Devices such as the Phipps Psychiatric Clinic Behavior Chart (4) have been used for a half century on psychiatric wards to record patient change. Plant (17) reported a rating scheme for describing patient behavior on the ward in 1922. In 1933, Moore (14) published his chart and "Schema for the Quantitative Measurement of Abnormal Emotional Conditions" containing some 36 carefully constructed scales. The interest in mental health problems intensified by events in World War II resulted in a marked upswing in the development of procedures for the objective measurement of psychopathology and personality change. It is the purpose of this review to examine briefly those rating scales and check lists designed to describe psychiatric patients on the ward or in the interview, which have appeared during the past ten years.

SCALES FOR USE BY PSYCHIATRIC AIDES AND NURSES

Six scales suitable for use in mental hospitals by nurses and psychiatric aides have been reported. These are generally designed to secure (a) quantified descriptive reports of readily observable patient ward behavior, and (b) quantitative estimates of hospital adjustment. The first of these, the Gardner Behavior Chart (22), a rating scale developed by Wilcox out of work with psychotic

patients, is designed to secure reports of easily observed patients' ward behavior from nurses and attendants. The 15 categories or scales used are: attention to personal appearance, sleep, appetite, sociability, activity control, noise disturbance control, temper control, combativeness control, care of property, self entertainment, cooperation in routine, work capacity, work initiative when alone, work initiative when closely supervised, and willingness to follow directions. Under each category five brief phrases characterize the grades of behavior in the scale. The rating grades are none, poor, fair, good, and extra good; they are weighted from 0 to 4. The total score consists of the sum of the 15 ratings received. The Behavior Chart has been found useful in the evaluation of change following prefrontal lobotomy (21).

The Fergus Falls Behavior Rating Sheet, prepared by Lucero and Meyer (10), was developed to record behavior of patients who are mute, unintelligible, hyperactive, or seclusive. Eleven aspects of behavior, such as work, response to meals, and response to patients, are rated by checking one of five descriptions. A value of 1 is given to the most deviant behavior and a 5 to presumed normal behavior. Thirty-four raters, on rating 51 patients, agreed 90 per cent of the time even though some of the language used in the description appears to be sufficiently ambiguous and difficult as to require extended training. A similar but briefer device, the Norwich Rating Scales, for

recording patient behavior on disturbed wards, has been developed by Cohen, Malmo, and Thale (2). The ward nurse or attendant checks one of five descriptions of activity, aggressiveness, destructiveness, resistiveness, talkativeness, and tidiness. To check the reliability of the individual scales, 10 patients were rated independently by two raters. The average interjudge coefficient reported was .76.

Rowell (18) has reported a graphic rating scale of 20 items for use by psychiatrically trained nurses. Individual scales are 5-point continua extending from normality on one end to pathology on the other. Variables such as preoccupation, hallucinations, delusions, affect, mood, blocking, and "flight of ideas," are rated. The reported immediate test-retest reliability for 71 total scores is .95, while the correlation between independent ratings by nurses for 62 pairs is .85. The descriptive statements preceding the scale cues appear far too terse and lacking in definition for terms such as blocking and flight of ideas that are ill-defined and notoriously difficult to judge.

The Hospital Adjustment Scale, an ingenious, carefully constructed device for evaluating patient's behavior, has been developed by Ferguson, McReynolds, and Ballachey (11). The scale consists of 91 statements descriptive of psychiatric patients, such as "the patient ignores the activities around him," or "the patient's talk is mostly not sensible." Each statement is marked as True, Not True, or Does Not Apply, for a given patient, and is keyed in such a manner that it is possible to obtain a total score indicative of the patient's general level of hospital adjustment. The scale can be filled out in about 10 minutes by the psychiatric aide or nurse most familiar

with the day-to-day behavior of the patient over a period of two weeks to three months. In addition to the total adjustment scores, the scale also offers measures descriptive of (a) communication and interpersonal relations, (b) care of self and social responsibility, and (c) work, recreation, and other activities. The statements can also be grouped as to whether they are indicative of an "expanding" personality, a "contracting" personality, or whether they are neutral. The Hospital Adjustment Scale was developed from a pool of statements descriptive of patients, secured from psychiatric aides. Statements were selected for the final form on the basis of measures of interjudge reliability, ratings made by 16 judges on a scale of over-all hospital adjustment, checks of discriminative power, and percentage of True, Not True, and Don't Know checks. Norms based on the records of 518 patients from four hospitals and clinics are available in percentile form. Patients approaching release from hospital can be differentiated significantly from those judged to be extremely disturbed or chronic hospital residents.

Scherer (20) has prepared a set of 44 four-point scales for the evaluation of patient behavior in such activities as occupational therapy, manual arts, corrective therapy, educational therapy, recreation, and library visits. The Activity Rating Scales were constructed on the basis of a longer list of patient behaviors and tested out on rehabilitation staff personnel. Independent ratings of the same group of patients indicated that, of 1,188 pairs of independent ratings, 60 per cent agree completely, 28 per cent differ by one scale interval, 3 per cent by two scale intervals, and 8 per cent of the scales are not rated. Inasmuch as the activities observed

in rehabilitative situations are similar to noninstitutional, vocational, and social activities, the author postulates that the behavior exhibited may also be similar to and predictive of behavior shown in posthospital adjustment.

SCALES FOR USE BY TRAINED CLINICIANS

The scales and check lists to be described in this section were developed primarily for use by trained clinicians. Most of them represent efforts to secure records of currently observable behavior, symptoms, complaints, or inferable needs and attitudes. A few, however, are intended to secure evaluations of social history obtained from the patient or the patient's friends and relatives.

The Elgin Prognostic Scale, constructed and validated by Wittman and Sternberg (28, 29), is a rating schedule designed to predict recovery in schizophrenics. It consists of 20 rating scales weighted according to prognostic importance; favorable factors are arbitrarily assigned negative weights and unfavorable factors are assigned positive weights. The prognostic score is the algebraic sum of the weighted measures. Most of the variables, such as shut-in personality, type of onset, or range of interests, are based upon premorbid social history secured from the patient's relatives or from the patient himself. A few are based on currently discernible symptomatology such as hebephrenic symptoms, ideas of influence, bizarre delusions, or affect. These scales were constructed following a review of available literature on prognostic factors in schizophrenia. They were validated and cross-validated on Elgin State Hospital patients, and shown to predict outcome of shock treatment with greater accuracy than staff judg-

ments. A multiple-factor analysis (6) of intercorrelations between the scales based on a group of 200 patients revealed three interpretable factors: schizoid withdrawal, schizophrenic reality distortion, and personality rigidity or inadaptability. The principal shortcoming of this device is the difficulty of securing reliable premorbid personality pictures from parents and relatives.

Saslow and his associates (19), in their effort to study the personality correlates of psychosomatic disorders, have prepared some 12 scales for measuring habitual patterns of reactions to crises. Each pattern is briefly delineated by means of a 5-point scale. Ratings are based on social history data secured from patients during psychiatric interview. Included are scales of impulsiveness, subnormal assertiveness, obsessive-compulsive behavior, depressive behavior, anxiety, hysteria, inward expression of emotions, low awareness of body symptoms, insecure feelings of inferiority, repressed hostility, and strong dependent needs. No reliability data are reported for the individual scales.

A check list resembling the Phipps Clinic Chart, but more systematic in its approach, has been developed by Peters (16). The check list, which consists of 199 traits, is grouped under 7 categories: history, acting, talking, mood, emotion, interests, ideation. The traits were compiled from interview records and presumably cover a large proportion of all characteristics required to describe a patient's personality and symptoms. Three ratings are used: a plus for a positive degree of a trait, no mark at all when a trait is present to a normal degree or does not pertain to the subject, and a minus sign for a negative degree of the trait. Ward behavior, interview data, and social history

may be taken into account in rating. The check list has been used successfully to identify traits related to improved adjustment of lobotomized patients (16).

A mental health check list entitled The Pattern for Living has been constructed and reported by Conrad (3). The check list is designed to measure positive mental health, social conformity, and pathology. It is for use by trained personnel in appraising persons applying for and receiving outpatient psychotherapy. Every item regarded as true is checked plus, false items are checked minus, while the remainder may be scored by a question mark. Trends are indicated in a separate column. Of the 45 items, 16 are concerned with positive mental health, 12 with conformity, and 17 with pathology. Evidence is presented that patients with high positive mental health scores tend to stay in therapy (3).

A Guided Clinical Interview Analysis for use in connection with structured clinical interviews has been reported by Abt (1). Eight scales descriptive of attitude toward parents, attitude toward siblings, attitude toward childhood, attitude toward people in general, and attitude toward sex make up the Analysis. Each subcategory of a scale is ratable on a 5-point scale. Agreement between raters for recorded interview material is reported to be high. A scale similar to the Abt form as to content, for use in evaluating adult outpatients receiving psychotherapy, is being developed by Morse (15). The six major areas measured are accessibility to therapy, occupational and school adjustment, social adjustment, sexual adjustment, family adjustment, and symptomatology. Most of the 40 items contain four brief descriptions to characterize grades of behavior on the scale con-

tinuum. Thus far neither reliability nor validity data are available.

The Psychiatric Rating Scales developed by Malamud *et al.* (13), at the Worcester Hospital in Massachusetts, represent yet another rating form for the quantitative recording of psychiatric clinical findings and the changes that occur in them during the course of illness. The scale consists of 19 "functions" divided into three major groups. The first seven functions comprise behavior items, such as sexuality and sleep, which depend upon continuous observations by ward personnel during the 24 hours preceding the psychiatric ratings. The third group consists of eight functions that are evaluated during the interview and require verbal reactions from the patients themselves. Associations, memory, and thought processes are examples of these eight functions.

Each of the 19 scales extends to the left and to the right from a central base line consisting of two terms which represent the usual variations in the particular function that is within normal limits. On both sides of the base line are indicated the range of deviation in terms of progressive degrees of pathology. On the left side of the scale are those deviations which are directed toward the outside (centrifugal). To the right of the base line are those centripetal or internally directed deviations in the function. Steps in the function are identified by three psychiatric terms to the left of the base line and three to the right. Ratings made for either half of a function may range from 1 to 6. The Feeling function, for example, is marked by the following sequence of terms: panic, anxiety-guilt, tense-irritable, hypersensitive, hyposensitive, phlegmatic, dull, apathetic. A patient's total score consists of the sum of his scores on each

of the 19 functions. A correlation of .92 was secured between 100 paired independent ratings made by two psychiatrists on 26 patients. Application of the Psychiatric Rating Scale to agitated depressives (12) indicates a good correlation with changes in the clinical picture resulting from electroshock therapy. The scale has also been used to record changes resulting from prefrontal lobotomy.

A criticism that may fairly be directed at the Psychiatric Rating Scales is that the terms used to describe the 19 functions are nowhere defined. This would suggest that ratings on individual functions are less reliable than the total scores. In total scores, differences between raters are cancelled out in the process of summing. Lockwood (5) reports that the evaluations of his psychiatric raters as reflected on individual scales were not sufficiently reliable for use as criteria for clinical improvement. In a number of the scales several variables seem to have been forced into a single bipolar continuum. Thus, instead of one rating there should be two on such functions as Sexuality. It would have been preferable to determine empirically in advance whether or not presumably antithetical forms of behavior were mutually exclusive.

The Elgin Behavior Rating Scale Revised has been developed by Wittman and Hills (29) for the purpose of describing psychiatric patients in six areas of behavior. A rating is accomplished by selecting the descriptive sentence listed under a category which is most applicable to the patient. The scales are graded from "very poor" to "very good." A weight of 0 is given to the most deviant behavior; intermediate steps are assigned weights from 1 to 4. Somatic behavior is covered by seven scales descriptive of physical appearance,

physical condition, appetite, and the like. The Social Behavior area consists of six scales descriptive of such aspects as conversation, cooperation, and sex behavior. There are eight scales to describe Mental Behavior such as orientation, insight, and affective response. Under the rubric of Psychotic and Neurotic Behavior are three global scales descriptive of affective exaggeration, paranoid projection, and schizoid withdrawal. Neurotic Behavior and Anti-Social Behavior are separately rated on two scales. No data with regard to inter-rater agreement are reported. Most of the scales contain several variables for rating. The orientation scale, for example, includes items relative to disorientation as to time, place, and person. There are actually 63 separate variables available for rating. Although no data are provided by the authors, the total score would probably represent a useful measure of over-all severity of illness.

A series of check lists designed to describe patient character, temperament, and intellectual capacity have also been developed by Wittman (29) and her associates. The check list of Fundamental Temperament Reactions postulate three bipolar components. The first component, affective exaggeration, extends from manic expansion to depressive constriction. Aggressive ascendance and defensive passivity mark the two ends of a continuum of paranoid compensation. The schizoid withdrawal component extends from heboid regression to simple withdrawal. Two descriptive elements identify manic expansion and a similar number are used to describe depressive constriction. The rater may check as few or as many elements as he observes; the total number of checks provides a total score. Wittman has also constructed check lists for Temperament

Deficiencies, Anxiety Reactions, Character Deficiencies, Exaggerated Reaction Types, Addictive Reaction Types, Disorders of Intellectual Capacity, Constitutional Intelligence Disorders (Amentia), and Acquired Intelligence Loss (Dementia).

Wittenborn (23) has reported a rating-scale procedure for the evaluation of mental hospital patients as one step in a broad program directed at the development of a quantified method for multiple psychiatric diagnoses. The procedure was devised to permit a psychologist, psychiatrist, nurse, or other competent observer to (a) rate the currently discernible symptoms of a psychiatric hospital patient, (b) score these ratings, and (c) prepare a profile which would indicate to what extent the patient's pattern of symptoms resembled the symptom patterns found among psychiatric hospital patients generally. The rating schedule consists of 55 different, unlabeled scales presented sequentially in a random manner. The scales were selected to represent a fairly adequate sample of important symptoms that characterize hospitalized patients. They were designed to demand a minimum of interpretation and experience from the observer, and to yield judgments which are relatively unbiased by the rater's particular theory or point of view. The directions require that the rater check the most pathological condition or level of behavior observed during the period studied and that every scale be checked for every patient. This last procedure undoubtedly facilitates correlational studies. However, there is a real question as to how much error is introduced if judgments are forced as in the case of the mute patient or the patient who is evasive or irrelevant in his speech. A mute patient, for example, must be judged as to rate of change of ideas, insight, or rate of speech.

Most of the scales consist of four elements; a lesser number contain three or five elements. The rating process, which consists of encircling one element for every scale, requires about 10 or 15 minutes.

In an effort to determine the existence of clusters on patterns of symptoms and the influence of such characteristics of the sample as age, sex, or organic brain damage on these patterns, Wittenborn (24, 25) has completed a series of factorial analyses of his rating schedule. The nine factors or psychiatric syndromes which have been repeatedly identified are: Acute Anxiety, Conversion Hysteria, Manic State, Depressed State, Schizophrenic Excitement, Paranoid Condition, Paranoid Schizophrenic, Hebephrenic Schizophrenic, and Phobic-Compulsive. Scoring weights for the symptom scales and norms for transmuting cluster scores into standard cluster scores are available for the Descriptive Scales for Rating Currently Discernible Psychopathology.

The Multidimensional Scale for Rating Psychiatric Patients (MSRPP) consists of two sets of brief, relatively objective rating scales for the description of the behavior, symptoms, complaints, and inferred motivation of psychiatric patients. One form or set is for outpatient use and the other for hospital use. These schedules represent a systematic effort to (a) develop a quantified record or description of mentally ill patients that could be used to measure change or improvement, and (b) isolate and identify underlying unitary variables. The form for outpatient use, developed by Lorr, Jenkins, Holsopple, and Rubinstein (8), consists of 49 unlabeled, randomly presented, 4- and 6-point graphic rating scales for describing outpatients as seen in diagnostic or therapeutic interview. It is intended only for use by psychologists or psy-

chiatrists. The scales, calling for judgments based on directly observable manifest behavior, are grouped together, as are those which require inferences. Included are scales descriptive of personality traits such as emotional responsiveness, complaints such as headaches, and symptoms such as compulsive behavior. The schedule provides scores for 16 factors identified on an original set of 73 (9). Tentative norms and standard scores for profiling individual patient records are available. The factors measured by this form have been labeled Hostility, Reality Distortion, Obsessive-Compulsive Reaction, Sex Conflict, Gastro-Intestinal Reaction, Cardiorespiratory Reaction, Anxiety-Tension, Anxious Depression, Emotional Responsiveness, Adaptability, Sense of Personal Adequacy, Vigorous Interest, Conscientiousness, Independent Maturity, Goal-Directed Motivation, and Prudence.

The MSRPP form for hospital use represents a revision of the Northport Record, which was initially constructed and developed by Lorr, Singer, and Zobel (7). It consists of 50 unlabeled graphic rating scales presented in a random order. The schedule was designed to measure the major symptoms characteristic of recognized syndromes of the various psychoses and behavior readily observable in a routine diagnostic interview or on the ward by nurses and psychiatric aides. A rating is made by encircling the entry which is most typical or representative of the patient during the period observed. When there is no basis for rating a patient on a particular scale, an unratable category is encircled.

The hospital form of the MSRPP provides tentative norms, based on 450 patients from four psychiatric hospitals, for 12 factors or syndromes measured. These factors were identified in a multiple-factor analysis of

55 of the original 81-item Northport Record. Standard score measures are available for profiling a patient on the following factors: Manic Excitement, Retarded Depression, Anxious Depression, Perceptual Distortion, Conceptual and Thinking Disorganization, Paranoid Suspicion, Grandiose Expansiveness, Schizophrenic Excitement, Disorientation, Withdrawal, Hostile Aggressiveness, and Activity Level. The last three of these factors are behavior parameters observed on the ward.

The sum of the absolute deviations from a "normal" pattern provides an over-all index of severity of illness which has been found to reflect change resulting from lobotomy. Seven type patterns similar to conventional diagnostic description are also available for use as an aid in diagnosis.

DISCUSSION

The development and use of rating scales and check lists for the systematic recording of clinical judgments of manifest behavior and inferred attitudes and needs appear to represent an important advance for the clinician and the research worker.

There seems to be no doubt that the interview is here to stay even though some critics, particularly the psychometricians, would have it replaced by other and presumably more rigorous procedures. The problem becomes one of developing controlled interview patterns as suggested by Zubin (31) and of objectively recording what the trained clinician can validly or reliably observe or infer. While the new techniques for sound recording of interviews are unquestionably important for evaluation, they are not substitutes for the clinician nor do they provide any complete basis for the analysis of the interview. Visual as well as auditory cues provide important data for an

appraisal. We are simply saying, in brief, that the clinician can contribute toward the description of his patient and the prediction of his future behavior. The rating form provides a framework for quantifying his judgments, for jogging his memory, and for minimizing "halo" bias.

The rating schedule also offers a common conceptual framework for the clinician regardless of the examination procedure used. Clinical judgments derived from an analysis of the Rorschach test, the TAT, or a sentence completion form may be recorded in objectified form on the rating scales. Ratings can be useful in defining and clarifying areas of agreement and disagreement. Clinicians differing in theoretical orientation can find a common ground when a concept characteristic of an individual is stated simply, in graded form. When defined in simple understandable terms, many presently elusive and amorphous variables can be checked for reliability and related to a larger domain of objectively expressed concepts. Conceptual formulations often loosely used, such as sexual identification or ego strength, can be pinned down for closer scrutiny and validation.

In any field where agreement on basic variables is lacking, factor analysis is a powerful tool for resolution of complex concepts into simpler elements and for the identification of underlying parameters. The confusion in, and duplication of, vocabularies for describing personality and psychodynamics are especially notable. However, Moore, Wittenborn, and others (14, 25) have shown that rating scales and check lists can be utilized fruitfully for the isolation and identification of psychopathological syndromes and categories. Investigators concerned with the isolation of primary factors in percep-

tion and cognition have recently recommended a battery of tests to represent each of the better defined factors in future factorial studies. In the absence of more objective measures of personality and psychopathology, it may be useful to utilize similarly, as reference variables, standard sets of rating scales in further factorial investigations.

SUMMARY

The purpose of this review has been to examine and report on rating scales and check lists designed to describe psychiatric patients in the interview and on the ward that have appeared during the past ten years. A half dozen scales and check lists suitable for use by nurses and psychiatric aides is reported in the literature. Of these devices the Hospital Adjustment Scale has been most carefully developed and seems to be usable by any psychiatric aide who can read. For more precise measurement of change the Gardner Behavior Chart and the Northampton Activity Rating Scale may be preferable.

Among those designed for use by psychologists and psychiatrists to describe psychotic behavior and symptomatology, the Wittenborn Descriptive Scales and the Multidimensional Scales have been most intensively analyzed and developed. The Elgin Prognostic Scales are at present the most useful for predicting improvement although difficult to use because of the practical problem of securing reliable data on the patient's past history.

The rating schedule offers considerable promise as a procedure for quantifying the interview, for isolating basic psychopathological parameters, and for generally providing a conceptual basis against which clinicians of varying persuasions and training can find a common ground.

REFERENCES

1. ABT, L. E. The analysis of structured clinical interviews. *J. clin. Psychol.*, 1949, **5**, 364.
2. COHEN, L. H., MALMO, R. B., & THALE, T. Measurement of chronic psychotic overactivity by the Norwich rating scale. *J. gen. Psychol.*, 1944, **30**, 65-74.
3. CONRAD, D. C. Towards a more productive concept of mental health. *Ment. Hyg.*, 1952, **36**, 456-473.
4. KEMPF, E. J. The behavior chart in mental diseases. *Amer. J. Insanity*, 1915, **71**, 761-772.
5. LOCKWOOD, W. L. Some relations between response to frustration (punishment) and outcome of electric convulsive therapy. *Comp. Psychol. Monogr.*, 1950, **20** (Ser. No. 104), 121-186.
6. LORR, M., WITTMAN, PHYLLIS, & SCHANBERGER, W. An analysis of the Elgin prognostic scale. *J. clin. Psychol.*, 1951, **7**, 260-262.
7. LORR, M., SINGER, M., & ZOBEL, H. Development of a record for the description of psychiatric patients. *Psychol. Serv. Cent. J.*, 1951, **3**, No. 3.
8. LORR, M., RUBINSTEIN, E., & JENKINS, R. L. A factor analysis of personality ratings of outpatients in psychotherapy. *J. abnorm. soc. Psychol.*, 1953, **48**, 511-514.
9. LORR, M., SCHAEFER, E., RUBINSTEIN, E. A., & JENKINS, R. L. An analysis of an outpatient rating scale. *J. clin. Psychol.*, 1953, **9**, 296-299.
10. LUCERO, R. J., & MEYER, B. T. A behavior rating scale suitable for use in mental hospitals. *J. clin. Psychol.*, 1951, **7**, 250-254.
11. McREYNOLDS, P., BALLACHEY, E. L., & FERGUSON, J. T. Development and evaluation of a behavioral scale for appraising the adjustment of hospitalized patients. *Amer. Psychologist*, 1952, **7**, 340. (Abstract)
12. MALAMUD, W., HOAGLAND, E., & KAUFMAN, I. C. A new psychiatric rating scale. *Psychosom. Med.*, 1946, **8**, 243-245.
13. MALAMUD, W., & SANDS, S. L. A revision of the psychiatric rating scale. *Amer. J. Psychiat.*, 1947, **104**, 231-237.
14. MOORE, T. V. The essential psychoses and their fundamental syndromes. *Studies in psychology and psychiatry*. III. Baltimore: Williams & Wilkins, 1933.
15. MORSE, P. W. Proposed technique for the evaluation of psychotherapy. *Amer. J. Orthopsychiat.*, in press.
16. PETERS, H. N. Traits related to improved adjustment of psychotics after lobotomy. *J. abnorm. soc. Psychol.*, 1947, **42**, 383-392.
17. PLANT, J. S. Rating scheme for conduct. *Amer. J. Psychiat.*, 1922, **1**, 547-572.
18. ROWELL, J. T. An objective method of evaluating mental status. *J. clin. Psychol.*, 1951, **7**, 255-259.
19. SASLOW, G., GRESSEL, G. C., SHOBE, F. O., DuBOIS, P. H., & SHROEDER, H. A. Possible etiologic relevance of personality factors in arterial hypertension. *Psychosom. Med.*, 1950, **12**, 293-302.
20. SCHERER, I. W. A behavior rating scale for use in activity therapy situations. *Info. Bull.*, Dep. Med. & Surg., Psychiat. & Neurol. Div., Veterans Administration, Jan., 1951.
21. SCHRADER, P. J., & ROBINSON, M. F. An evaluation of prefrontal lobotomy through ward behavior. *J. abnorm. soc. Psychol.*, 1945, **40**, 61-69.
22. WILCOX, P. H. The Gardner behavior chart. *Amer. J. Psychiat.*, 1942, **98**, 874-880.
23. WITTENBORN, J. R. A new procedure for evaluating mental hospital patients. *J. consult. Psychol.*, 1950, **14**, 500-501.
24. WITTENBORN, J. R. Symptom patterns in a group of mental hospital patients. *J. consult. Psychol.*, 1951, **15**, 290-302.
25. WITTENBORN, J. R., & HOLZBERG, J. D. The generality of psychiatric syndromes. *J. consult. Psychol.*, 1951, **15**, 372-380.
26. WITTENBORN, J. R., MANDLER, G., & WATERHOUSE, I. K. Symptom patterns in youthful mental hospital patients. *J. clin. Psychol.*, 1951, **7**, 323-327.
27. WITTENBORN, J. R., BELL, E. G., & LESSER, G. S. Symptom patterns among organic patients of advanced age. *J. clin. Psychol.*, 1951, **7**, 328-331.
28. WITTMAN, PHYLLIS. A scale for measuring prognosis in schizophrenic patients. *Elgin Pap.*, 1941, **4**, 20-33.
29. WITTMAN, PHYLLIS. The Elgin check list of fundamental psychotic behavior reactions. *Amer. Psychologist*, 1948, **3**, 280. (Abstract)
30. WITTMAN, PHYLLIS, & STERNBERG, L. Follow-up of an objective evaluation of prognosis in dementia praecox and manic-depressive psychoses. *Elgin Pap.*, 1944, **5**, 216-227.
31. ZUBIN, J. Objective evaluation of personality tests. *Amer. J. Psychiat.*, 1951, **107**, 569-576.

Received April 9, 1953.

RECENT STUDIES OF SIMPLE REACTION TIME^{1,2}

WARREN H. TEICHNER

Aero-Medical Laboratory, Wright-Patterson Air Force Base

In spite of the important role that the human reaction time study has played in the development of psychological science and the tremendous amount of research effort expended in its behalf, there are still large gaps in our knowledge of the empirical relationships in which reaction time is involved. This report is an assessment of the present scientific status of the topic based primarily on the experimental literature of the last twenty years. Previous reviews were presented by Woodworth (161) in 1938 and by Johnson (82) in 1923. In addition, Forlano, Barmack, and Coakley (50) have reviewed the effects of ambient and body temperature on both simple and choice reaction time and Finan, Finan, and Hartson (46) have briefly summarized the use of reaction time scores as measures of performance decrement.

First, it should be noted that there are several kinds of reaction time experiments (8, 161). Because of the

tremendous literature involved, the present discussion will be restricted to the *simple reaction time* (RT), which is the time interval between the onset of the stimulus and the initiation of the response under the condition that *S* has been instructed to respond as rapidly as possible. In order of presentation this review will consider the effects on the RT of stimulus and receptor conditions, central and motor factors, and certain special conditions such as the effects of low ambient temperature, loss of sleep, etc.

Next, the complexity of the measure should be recognized. After the onset of the stimulus there is a lag, or latent period, during which the receptor process is initiated and builds up to a maximum (25, 56, 119, 120, 140, 141). This is followed by a second lag involving central transmission of the sensory impulses to the motor fibers, and finally, there is a time delay involved in the contraction of the muscles (33, 34, 47, 67, 127) and the beginning of the movement of the responding member. Any of the factors which affect any of these processes will obviously also affect the measured RT within which they are present. Since most RT studies deal only with the over-all measure of time, the present discussion will be confined to this measure. However, a discussion of sensory latent periods relevant to the visual RT may be found in Arnold and Tinker (2) and in Strughold (140, 141), to the auditory RT in Chochell (24, 25), to the pain RT in Pattle and

¹ This paper is a revision and extension of *The Simple Reaction Time, a Review with Reference to Air Force Equipment*, *Wright Air Development Center Technical Note WCRD 52-47*, August 1952, which was written when the author was on the staff of the Psychology Branch, Aero-Medical Laboratory, Wright Air Development Center. The writer is now with the Human Resources Branch, Natick QM Research & Development Laboratory, Natick, Massachusetts.

² Thanks are due to many people for their comments, and in particular to Mr. Darwin Hunt and Dr. Davis Howes of the Aero-Medical Laboratory, and Dr. Austin Henschel of the Natick QM Research and Development Laboratory.

Weddell (117), and to the thermal RT in Wright (162). Piéron (120) summarizes the topic in considerable detail.

STIMULUS-RECEPTOR FACTORS

RT as a Function of the Sense Modality Stimulated

It is a common assumption based on the neuro-anatomical differences existing among the various receptor systems that the time of reaction varies according to the sense modality stimulated. Textbooks usually contain comparisons between the RT's obtained by various kinds of stimulation, frequently presenting lists in which RT is ranked according to the senses. However, little attention appears to have been given to the logic of measurement involved in making such comparisons. For example, in order to say that the auditory RT is shorter than the visual one, as is usually done, the two types of stimulation must be compared on the same scale. The scales that have been employed are, unfortunately, scales of subjective intensity and these cannot be considered comparable from one sense modality to another. This argument, i.e., that the attribute terms of one psychophysical dimension logically cannot be projected to another such dimension, represents what is usually thought of as an advance in the logical foundation of psychology and will be remembered as having invoked considerable discussion.³ Although the problem was resolved with regard to the comparing of sensations, it seems to have been ignored in dealing with the RT. For this reason, the conclusion that must be drawn is that *there is no evidence available that indicates whether*

or not the RT varies according to the receptor system stimulated.

The experimental literature available does allow one kind of meaningful conclusion regarding this matter. Where studies have been done comparing a specific intensity of, say, sound with a specific intensity of, say, light, it should be possible to decide whether the RT's for those specific values, and those only, are shorter for the one sense than for the other. Unfortunately, most studies making such comparisons either report no intensity values at all, or they report them in arbitrary units with no means of reference.

It is possible, of course, to speculate. The literature concerned with visual and auditory RT's is almost universal in reporting faster RT's to the sound stimuli than to other stimuli. Most studies have also found faster RT's to tactual than to visual stimuli. Robinson (125), for example, presented a summary table of eight of the older investigations in which the RT's for vision, hearing, and touch were compared. If medians are calculated from this table, the RT for audition is found to be 0.142 sec., for touch 0.155 sec., and for vision 0.194 sec. In all eight of the experiments, the auditory RT was consistent in being faster than the visual one. But in four of the eight, the tactual RT was faster than the auditory one and in the other four the opposite result was obtained. Todd (146), and in addition more recent studies (4, 16, 38, 48, 153, 154), concur in finding shorter RT's to sound than to light. One other study (Moldenhauer) reviewed by Woodworth (161) found the auditory RT faster than the tactual one. Lanier (92), however, in a study of the effect of training on auditory, visual, and tactual RT's found the

³ For a complete statement of this problem see Boring (10).

tactual RT shortest for trained Ss with the auditory and visual RT's being approximately equal. With untrained Ss, on the other hand, the auditory RT was shortest with visual and tactual times being equal.

It is possible that the speed of reaction with respect to sensory modality may depend on some sort of speed or reaction factor which determines the kind of stimulus to which an individual will exhibit the shortest RT. Wells, Kelley, and Murphy (153, 154) studied the ratio of the median RT to light to the median RT to sound and found a light-sound ratio of 1.15 for 11 untrained Ss. Two trained Ss exhibited ratios of 1.34 and 1.45. They also found a correlation of $-.52$ between the ratios and the median RT to sound in the untrained group. From this they concluded that Ss who have a relatively fast RT to sound have a relatively slow RT to light, and vice versa. The ratios obtained agree fairly well with those recently reported by Canfield, Comrey, and Wilson (16). However, the correlation of $-.52$ has little meaning since it is logically possible to obtain a negative correlation between a ratio and its denominator even when the direct correlation between the numerator and the denominator is positive. In fact a negative correlation merely indicates that the ratio is greater than 1.00. More direct comparisons by Forbes (48) and by Lanier (92) have revealed positive correlations of .48 and .90 respectively between response to light and to sound.

From a consideration of chemical, temperature, pressure, or electrical stimuli applied directly to the skin to elicit pain response, Woodworth (161) concluded that the slowest RT is that based on painful stimulation. Recent studies have also recorded pain RT's to radiant heat (e.g., 61, 139, 156)

and radiant cold (35). It is necessary, however, to consider not only the measurement problems involved but also the logical and technical difficulties involved in the measurement of pain (37, 61).

Wright (162) investigated the RT's associated with cutaneous sensations of warmth elicited by visible and infrared radiation. In 22 out of 39 men tested, RT was faster to stimulation on the back of the hand than it was on the palm. In 17 out of 27 others who were tested, stimulation in the epigastrium produced faster RT's than stimulation in the interscapular region. In 78 Ss the RT to light was found to vary from about 0.33 sec. for a very intense sensation to about 20 sec. at threshold. One interesting result was that psychophysical functions obtained with RT measures showed a qualitative similarity to visual functions of intensity, duration, and area.

A little work has also been done recently on proprioceptive RT's. Chernikoff and Taylor (21) measured the speed of reaction to the sudden falling of the S's arm. When the response measure was the time of release of a telegraph key, no differences were obtained between auditory, tactual, and kinesthetic RT's. However, when the response was the stopping of the arm movement, RT was considerably shorter than the key-release type of reaction regardless of whether the latter was to an auditory, tactual, or kinesthetic stimulus. Hick (72) and Vince (152) have also studied the RT's involved in making corrective movements in a pursuit task involving both visual and proprioceptive components.

In a different type of proprioceptive study, Baxter and Travis (5) measured the vestibular RT's of 31 Ss to rotary motion of the body. The Ss were blindfolded, seated in a con-

stant-speed revolving chair, and instructed to press a telegraph key upon detecting a change in motion. When the chair was moved from a stationary position, a mean RT of 0.52 sec. was obtained; when the chair was in motion and the direction of motion was changed, a mean RT of 0.72 sec. was obtained. From this highly significant difference Baxter and Travis concluded that the RT to perception of motion from rest is faster than the RT to perception of change in direction of motion.

The possible variation of RT with the sense modality used for stimulation is a question which has not been answered by the studies reviewed. As noted above, genuine comparisons require the use of common scales and such scales have not been employed. One possibility presently available is to determine the RT's on a statistical or probability basis and make comparisons in terms of a probability of RT scale. The first step is to determine empirically for each sensory modality the relationship between RT and the intensity dimension used for that modality, i.e., the function $P(RT)=f(I)$. This relationship is presumably sigmoid. Time of reaction (RT) can now be determined as a function of the probability of reaction for each case, i.e., the function, $RT=g[P(RT)]$ can now be obtained. Since $P(RT)$ is common for all modalities, this function furnishes a legitimate basis of comparison within the range, $0 < P < 1.00$.

RT as a Function of the Number of Sense Organs Stimulated

No recent work has been done on mono- versus bisensory stimulation. As indicated in previous reviews Poffenberger (121) reported the RT to light about 0.015 sec. faster for each of the three Ss he used when both eyes were stimulated than when only

one was stimulated. This is hardly conclusive considering the size of his sample and the likely error of measurement. Similarly, Bliss (9) found the RT slightly shorter for binaural than for monaural stimulation. More recently, Smith (136) found faster RT's to the apparent movement of objects when viewed binocularly than when viewed monocularly.

Since both visual and auditory phenomena usually show differences according to whether the stimulation is mono- or bisensory, it is reasonable that RT's should show corresponding differences. The high expectation of these differences is probably the reason why experimenters have been so little motivated to provide demonstrations of them. This is unfortunate because neither Poffenberger's data nor those of Bliss should be considered reliable enough to be sufficient.

Simultaneous presentation of stimuli to different senses has also been studied. Todd (146) presented light, sound, and electric shock singly and in simultaneous combinations and measured the speed of simple reaction to each. Combined stimuli in every case elicited faster RT's than the individual stimuli making up the combination. The RT to combined sound and light, for example, was not only faster than to light alone but also faster than to sound alone. The shortest RT's were to the combination of all three types of stimulus. On the other hand, *successive* stimulation of different sense organs produced longer RT's than did stimulation of single sense organs.

RT as a Function of Number of Receptors Stimulated

According to neurological principles of summation, it would be expected that the greater the number of receptors stimulated, the shorter the

latent period and, consequently, the shorter the time of reaction. There are two studies which support this hypothesis. Older data from Froeberg (55) provide some support in the visual case. These data show that, as the retinal area stimulated is increased from three to 48 sq. mm., RT decreases from 0.195 sec. to 0.180 sec., the function appearing to be negatively accelerated and still decreasing at 48 sq. mm. Wright (162) reported a similar type of phenomenon for thermal RT's.

RT as a Function of the Location of the Stimulation in the Visual Field

The visual RT varies with the portion of the visual field stimulated, according to Poffenberger (121). This investigator stimulated at 3°, 10°, 30°, and 45° away from the fovea and measured the increase in length of RT over the RT obtained from foveal stimulation. He found that RT increased in the temporal periphery from about 0.004 sec. at 3° to about 0.024 sec. at 45°, and in the nasal periphery from about 0.004 sec. to about 0.015 sec. Except for stimulation at 3°, increases in RT were consistently greater for temporal as compared to nasal stimulation.

Poffenberger's data suggest that RT is positively correlated with ability to perceive shape since shape perception is greatest in the foveal area and decreases toward the periphery. RT seems to be correlated also with visual acuity (reading of numbers, letters, etc.) since this too decreases with distance away from the fovea and is greater in the nasal half. These relationships, although not recently confirmed, present some very interesting possibilities for the field of visual measurement and suggest, among other things, speed-of-seeing techniques for the testing of visual acuity. Such techniques have had

some use in the measurement of acuity as related to the amount of illumination (26, 45, 96).

Longer RT's with stimulation in the peripheral area would not be expected in the dark-adapted eye, peripheral sensitivity to weak light being greater under this condition. Data relevant to this last hypothesis were obtained by Lemmon and Geisinger (94) who measured the RT's of 14 Ss at the fovea and 45° away. Under light-adapted conditions they found the RT significantly longer in the periphery, which supports Poffenberger's (121) results. When the eye was dark-adapted, they found the average RT of their Ss slightly shorter in the periphery, which is in accord with the hypothesis. This result, however, was not statistically significant.

RT as a Function of the Intensity of the Stimulus

Considerable research and theoretical effort have been expended on the relationship between RT and stimulus intensity. Early studies (7, 20, 26, 45, 55, 56) all agree that the visual RT becomes shorter as the intensity of light is increased. More recent investigations (32, 96, 137, 138) concur. In spite of earlier controversies, there is little doubt that the relationship is a nonlinear one, not only in the case of the visual RT but also for auditory (24, 25, 118), gustatory (118), thermal (162), and pain (35, 73, 120) RT's. Attempts have been made to fit the intensity data into mathematical, theoretical frameworks, with exponential, hyperbolic, and parabolic functions all being used more or less successfully on the same sets of data (78, 91, 118, 124).

In all the intensity studies cited, the intensity of the stimulus was varied and the speed of reaction measured when the receptor was in a "normal"

condition. In a related, but somewhat different type of study, Hovland (76) carefully investigated the effect on the RT to a test stimulus of 250 foot-candles of having the eye previously adapted to lights ranging from zero to 200 foot-candles. As might be expected, RT became shorter as the difference between adaptation brightness and stimulus brightness increased. This result is not altogether clear, however, since Lemmon and Geisinger (94), who also used a very bright test stimulus, found shorter RT's with the light-adapted than with the dark-adapted eye.

In addition to varying the intensity of the stimulus or of the adapting stimulus, it is also possible to vary the magnitude of change in the intensity of on-going stimulation. The RT has been studied in this way only in vision. Steinman (137) found that the RT became shorter as the relative magnitude of the change in intensity was increased. The relationship held only up to a limit, however, after which the RT began to increase again. This suggests that the function is not monotonic, but rather has an optimum at some moderate amount of change in intensity. If this conclusion is supported, it may provide an explanation for the discrepancy between Hovland's (76) finding a consistent decrease of visual RT with greater differences between adaptation intensity and test stimulus intensity and the opposing result obtained by Lemmon and Geisinger (94); i.e., the change in stimulus intensity in the latter study may have been at or beyond that point in the function where RT begins to become longer. Some support for this hypothesis is to be found in a study reported by Johnson (81). In this experiment Johnson darkened one half of a photometric field by about 4.6 per cent and compared the RT's

when the surround was 2.25 times as bright as the test field, 0.75 times as bright as the test field, or too dark to be measured. The test field itself had a brightness of 7.8 millilamberts and was of foveal dimensions. Johnson reports that the slowest RT was that elicited under the darkest condition, the next slowest under the brightest condition, and the fastest RT was obtained under the moderately bright surround. All differences were statistically significant. These results, along with those of Steinman, indicate the likelihood of an optimum intensity change and again suggest an explanation for the otherwise contradictory results of Lemmon and Geisinger.

RT's have been used effectively to study the effect of illumination on visual acuity. Luckiesh (96, p. 131) and Cobb (26) both found that speed of vision (RT) increases rapidly with increases in illumination up to about 18 to 20 foot-lamberts, after which further increases in illumination have no significant effect. Data from Ferree and Rand (45) indicate that speed of seeing is also a function of size of test object. In this study the limit of speed reached a maximum at approximately 25 foot-candles for relatively large test objects (3, 4.2, and 5.2 minutes of visual angle) and at approximately 45 foot-candles for relatively small test objects (1 and 2 minutes of visual angle).

RT as an Index or Measure of Sensation

The study of the effect of changes in magnitude of stimulus intensity on RT suggests the possibility of applying RT measures to psychophysical problems. The use of choice reaction times in what has been called the method of judgment time (17, 40, 144) is not new, of course. Although the use of the RT as a psychophysical

measure has always been implicit in the design of experiments involving visual and auditory thresholds, it has only recently had serious employment in this regard.

Steinman (137) studied the adequacy of the RT to change in brightness as a psychophysical method. As discussed above, he found that RT decreased as a (seemingly hyperbolic) function of the magnitude of change. With a constant stimulus-ratio this relationship was maintained up to the higher intensity levels where a reversal occurred. Although he attributed this reversal to an adaptation effect, other hypotheses are possible, as was indicated in the discussion of the effect of intensity.

Steinman (137, 138) also observed that the RT was faster to a decremental change in magnitude than it was for an objectively equal increment of change. This places a restriction on the method, but it is a restriction not without parallel in the standard psychophysical techniques. In any case, Steinman was able to plot RT functions which were in close agreement with similar functions obtained by more customary procedures, and for this reason he concluded that the method is adequate for securing equal perceptibility contours, threshold measurements, etc.

Other similar studies have been done, most of them quite recently. Galifret and Piéron (57) were also successful in obtaining visual functions based on differences among RT's. Chocholle (25) discusses the problem relevant to the psychophysiology of hearing; Wertheimer (156) suggests using the RT for obtaining radiant heat pain thresholds. Essentially this has been done both for the determination of the radiant heat pain threshold (61, 139) and the radiant cold pain threshold (35). Wright (162), furthermore, has obtained the

Weber type of function by using cutaneous RT's as an index of sensations of warmth produced by light radiation.

RT as a Function of the Duration of the Stimulus

It is difficult to see why the duration of the stimulus should influence the RT to the onset of a suprathreshold stimulus unless some type of summation of intensity hypothesis is advanced. Nevertheless, there is some suggestion in the literature that stimulus duration does have an effect.

Froeberg (55) varied visual stimuli by equal geometric intervals between 0.003 sec. and 0.048 sec. Within this range he found that the longest durations produced the shortest RT's, the function (of the geometric intervals) being linear.

Wells (155) varied the duration both of sound and of light stimuli. For an auditory stimulus he used the sound of an electric buzzer with durations of 0.007, 0.036, 0.051, 0.076, and 0.106 sec. Two Ss gave 200 responses under each duration. The results indicated that RT was a linear function of the logarithm of the duration of the stimulus, which is in agreement with Froeberg's visual data. But, although the form of the relationship was the same as that obtained by Froeberg, *the slope was in the opposite direction*. To study the effect of duration in the visual case, Wells used a constant intensity stimulus (brightness of 0.12 millilamberts) at five durations ranging between 0.012 and 1.00 sec. In this experiment Ss responded to the onset of the light. In a second experiment performed with the same Ss, response was made to the cessation of the light. Five durations of light were used again, this time ranging between 0.010 and 1.00 sec. The results differed from those of Froeberg in that they indi-

cated that there is an optimal duration and that this duration varies from individual to individual. Whatever the individual optimum, RT tended to become longer with deviations from it. The range of optima for ten Ss was between 0.025 and 0.066 sec. However, large variations were found not only for the optima among Ss, but in the optima between stimulus onset and cessation even for individual Ss; i.e., the optimum was usually not nearly the same for a single S under the two conditions.

Brogden and his students (21, 22, 62, 63) recently performed a series of experiments in which they compared the RT to auditory stimuli of fixed duration with the RT for response-terminated durations. They concluded that the primary variable is stimulus duration and that response termination merely acts on this factor. However, they also found that for longer sound durations (400–2000 msec.) the RT to the response-terminated stimulus was slightly, but significantly, shorter than to the fixed duration stimulus. For the shorter durations (100–200 msec.) there were no differences. Scrutiny of their data leaves the impression that RT increased in general as stimulus duration was increased from 100 msec. to 400 msec., and as the duration became still longer (up to 2000 msec.) RT decreased again. This is in agreement with Wells's results with visual stimuli, but not with Wells's auditory data or with Froeborg's visual data.

In spite of the conflicting results obtained, it does seem as though there is a relationship between stimulus duration and RT. The most reasonable expectation is that the function is asymmetrical in nature, falling (RT becoming shorter) rapidly as the duration is lengthened from zero to some small time value, whereupon it rises more gradually to

become asymptotic to some limit. The minimum of this curve is itself probably some function of stimulus intensity and length of foreperiod.

RT as a Function of the Onset and of the Cessation of the Stimulus

Another feature of the stimulus presentation which has occasioned some research, but about which little conclusive can be said, is the use of the onset and of the cessation of the stimulus as the signal for response. Both Holmes (74) and Jenkins (79) report shorter RT's to the cessation of light. On the other hand, Woodrow (160) found no differences in speed of simple reaction to stimulus onset or cessation for either auditory or visual stimuli. Wells (155) has shown that individual differences are very marked here, some Ss reacting more quickly to one or the other. Woodworth (161) concludes that there is no difference between the two conditions, but the matter seems far from settled. Studies are required which control the subjective intensity of the stimulation at the time of onset and of cessation. The duration of the stimulus would appear to be an important variable differentially determining the effectiveness of each condition. The readiness of S might be expected to be different in accordance with the one to which he responds.

CENTRAL AND MOTOR FACTORS

Cerebral factors have not been studied in a way to make them applicable to this discussion, although a little work has been done (135). Studies of speed of nerve conduction have theoretical significance but do not allow for generalization since they deal with the application of the stimulus directly to the nerve fiber and by-pass both sensory and motor factors. Certain psychological condi-

tions which might be of importance, e.g., personality factors (12, 151), the influence of incentives and punishments (80), etc., will not be discussed owing to the limited research that has been done and the great complexities of the results. A few topics that might have been included here have been singled out for the next section, Special Factors.

Regarding the motor system, most of the work that has been done on the responding mechanism is irrelevant in this context. Studies of the refractory phase (e.g., 142) are of interest only in the case of successive quickly elicited responses (39, 123). Most of the relevant studies of this sort involve more complex types of reaction time measures.

RT as a Function of Age and Sex

The correlation of age and RT has received some attention. Jones (84) found an increase in speed of reaction to sound in boys ranging from 11 to 14 years. According to this study no further increases should be expected beyond age 14. Atwell and Elbel (3), however, report continued small increases in speed up to 17, the oldest age used in their study. Bellis (6), in a study of ages ranging from 4 to 60 years, observed a general shortening of both visual and auditory RT's until age 30, after which latencies began to grow longer. Even at 60, however, he found that RT is still faster than it is at 10. Miles (107), using three kinds of response (finger pressing, finger lifting, and right-foot lifting), tested 100 adults between 25 and 87 years of age and found low (0.25 to 0.55) but significant positive correlations between age and speed of reaction. Elliot and Louttit (38) also report a low positive correlation. It would appear, therefore, that RT does covary with age. "Years old,"

however, may not be the best measure of the age factor.

Regarding sex differences, Elliot and Louttit (38) in an investigation of the braking reaction of men and women in automobiles reported that men react significantly more quickly. Bellis' (6) data, based on both auditory and visual stimuli, also favor males, especially in the age periods of 4-10 and 40-60. Seashore and Seashore (132), in studying the speed of various muscular responses (right and left hands, right and left feet, jaws), found men significantly faster, especially after practice. The weight of the evidence indicates that a sex difference favoring men does exist, although this conclusion is hardly likely to end a perennial controversy.

RT as a Function of Preparatory Set

It is reasonable that RT will depend on the degree to which *S* is ready to respond. The use of a preparatory signal has been shown to yield faster RT's than the omission of such a signal (163) or the unexpected presentation of a second signal requiring a second reaction (123). Consequently, most experiments use a ready signal.

This factor of readiness or *set* seems to depend, among other things, on the length of time between the warning or ready signal and the stimulus to which the response is made, what is known as the *foreperiod of reaction*. Breitweiser (11) found definite individual differences in the length of the optimum foreperiod and reported a range of optima between 1.0 and 4.0 sec. Telford (142), however, taking repeated measurements of 29 Ss, obtained conflicting results when he found that the average RT increased systematically from an optimum of 1.0 sec. to at least 4.0 sec. Telford's data also indicated that as the fore-

period is reduced from the 1.0 sec. optimum to at least 0.5 sec., RT deteriorates markedly.

The study most frequently quoted with regard to the foreperiod is that of Woodrow (159). According to this study a 2.0-sec. foreperiod is optimum. In spite of the wide acceptance of this figure, there is a question as to the significance of Woodrow's results since the data were obtained from only three Ss and, as they are reported, do not allow for an estimation of the standard errors of the means. Freeman and Kendall (54) estimated that if Woodrow's standard errors (which were not reported) were of the same order as those obtained in their study, which used four Ss, the difference between the 2-sec. and 8-sec. intervals obtained by Woodrow would have fallen between the 5 and 10 per cent levels of confidence. Ordinarily this would be sufficient to accept the null hypothesis.

Early studies (85, 158) demonstrated that preparatory sets are primarily muscular. Livingston (95) has shown that the amount of muscular tension varies during the foreperiod. Freeman (52, 53) and Freeman and Kendall (54) have investigated the influence of the amount, the locus, and the time of induction of muscular tension during the foreperiod on the RT. Under conditions of heavy load (large induced muscular tension), a longer foreperiod was found to be optimum than under light-load conditions. The locus of the tension was also found to be a significant factor in determining the optimum preparatory interval. Perhaps their most interesting finding was that the optimum interval varied with the length of time prior to the response that the tension was induced. When the amount, locus, and time of induction of muscular tension

are all considered together, they (54) found that the optimum foreperiod ranged between 4 and 8 sec., depending on individual differences. As implied above, there is reason to believe that Woodrow's obtained optimum actually may be expressed best as a range between 2 and 8 sec.

Another factor thought to influence the readiness of the S to respond is whether he is set to respond to the stimulus ("sensory attitude") or whether he concentrates on the response ("muscular attitude"). Although the weight of the evidence appears to favor the latter, no controlled studies have been performed since Woodworth's (161) review. We are forced, therefore, to rest with the older, and for the most part less well-controlled, studies. These are thoroughly discussed by Woodworth.

The effect of instructions constitutes another great unknown which presumably determines the readiness of the S to respond. Perhaps the most serious obstacle to overcome in investigations of this factor is not the ambiguity of language but an inability to know and to control S's self-instructions. Some promising work has been done, however, in which instructions were varied. Davis (33) compared the effect of instructions to respond with instructions to *not-respond*. Instructions to respond resulted in higher prestimulus tension levels in the responding arm, as measured by action potentials, and resulted in faster RT's. Davis also noted that instructions to respond were made even more effective by using more intense stimulation. This, however, would be predicted from the relationship between RT and stimulus intensity and may, therefore, be independent of the effects of instructions.

Moore (111) found that instruc-

tions have a greater influence on the variability of the RT than they do on the variability of the speed of movement of the responding part. When Ss were instructed to respond as quickly as possible with the fastest possible movement, the speed of movement remained practically constant but the speed of reaction varied considerably.

It is clear that a great many factors influence the optimum foreperiod. Other variables which should be studied in this relation are the intensity and duration of the stimulus. Of these, duration would appear especially important since the durations of both ready signal and stimulus are confounded with the length of the foreperiod. No single value seems acceptable as *the optimum* since so many conditions are effective. Perhaps a more useful concept would be a range, the modal point of which would shift according to sensory modality, stimulus intensity and duration, nature of muscular tension, kind of instructions, etc. Presently available data indicate that such a range lies between approximately 1.5 and 8 sec.

RT as a Function of Body Position

One question of much interest is whether or not the RT varies according to the position of the body. This has especial relevance today in connection with many military and industrial problems. Reference to one such study has been found. Munnich (114) investigated the RT of several responses modeled after various aspects of an airplane pilot's task. These RT's were studied with S in six different bodily positions: (a) seated normally with the back of the chair making a 120° angle with the seat; (b) stomach downward; (c) head downward; (d) back downward; (e) on right side; (f) on left side. Un-

fortunately, the report was not available at the time of writing and the abstract does not describe the differences, if any, which were obtained between positions. One interesting result, however, is described. It was found that RT increased in general after any *change* in position, but as the new position was maintained the RT's tended to return to the values normal for it. The generality of this conclusion is offset by the finding that some Ss improved their speed of reaction even after being changed to the most uncomfortable positions.

RT as a Function of the Responding Member

Of great theoretical and applied importance is the question of whether the RT varies with the body member which responds and with the type of motion involved. Regarding the latter, most of the responses which have been used have consisted of the simple release of a telegraph or similar key, or, less frequently, the depression of such a key. Other types of hand motion have also been used and, in addition, such other body parts as the eye, mouth (both movement of the jaws and the verbal response), arm, leg, foot, or the entire body. Although the types of response are usually as simple as possible and the methods of measurement usually some sort of chronoscopic device, both of these vary somewhat from one experiment to another. Discussions of the experimental techniques used and of the effects of these techniques may be found in several places (8, 11, 161). Whether the equivocal results obtained from studies of this type are really due to differences in technique is hard to say.

Woodworth (161), in reviewing this problem, presented a number of studies in which no differences were found between body members.

Baxter's (4) data support these results. On the other hand, Féré (44) found that the RT for the left hand was slower than that for the right, and that the sum of the RT's of each hand was larger when *S* tried to respond with both hands at once than when he moved them successively. Cattell (20) also observed that the RT for a particular task differs with the body member used, at least for the wrist, forearm, and shoulder. Hathaway (66) obtained results indicating that RT is longer for movement of the entire arm than it is for finger movement. Seashore and Seashore (132) reported that RT with the left hand was slower than with the right, and, similarly, the left foot was slower than the right foot. They also found that speed of reaction of the jaws was greater than either the hands or feet. Chernikoff and Taylor (23) presented data showing that when the stimulus is the sudden displacement of *S*'s arm, his RT is shorter when he stops the movement than when he releases a telegraph key. Considering the contradictions in the various sets of data, it certainly seems too early to draw even a tentative conclusion about the role played, if any, by the kind of responding member. It is possible that the only real differences are due to differences in inertia of the responding musculatures.

Woodworth (161) suggested that complex movements produce longer RT's than do more simple ones. Not many studies were available to support this, but on the basis of those few that were, Woodworth made two specific hypotheses: (a) that guided (aimed) movements have longer RT's than those made freely; (b) that the RT is shorter for the initiation of a movement than it is for the stopping of it or for the changing of its direction.

No studies were found bearing directly on either of these hypotheses, but there are some results which are relevant to the general hypothesis that RT is somehow related to the response movement that follows it. Searle and Taylor (129) investigated the RT of corrective tracking movements made with a small knob to the displacement of a visual stimulus and found that this RT was not affected by the amount of knob movement required to make a correction. Brown and Slater-Hammel (13) reported that the RT involved in making discrete arm movements in the horizontal plane is independent of the length and direction of the movement. Henry (71) obtained data which indicate that RT's and movement times are not related. Except for the possible effect of the preparatory set, it is difficult to see why RT and consequent movement should be related. The results available indicate that they are not.

Other Central-Motor Factors

Studies of extraneous tension during the occurrence of the response appear to be of only incidental interest here. Action potentials taken from muscles not involved in the response are not related to the speed of reaction, according to Meyer (105, 106). Daniel (30), however, considered Meyer's conclusion not warranted by his data, and Henderson (69) reported a relationship between speed of reaction of one arm and action potentials taken from the non-participating arm.

Of greater interest in this context is the finding that lowered muscular tension produces longer RT's (86, 87, 147). This finding appears to be related to the problem of the preparatory set and to the effects of practice. A few writers (e.g., 69) have reported increased speeds with practice. It is

possible that these increases may be due, not to the effect of learning on the RT itself, but to the effect of learning on the preparatory interval. This latter effect may consist of the learning of an optimal anticipatory muscular tension.

The RT exhibits considerable variability among individuals (41, 60, 92, 161) but is, nevertheless, reasonably consistent. Reported reliability coefficients range between 0.83 and 0.92 (41, 42, 59, 115, 130, 131). RT measures also show considerable variability among experimental studies. Both types of variability may be due, in part, to the points on the tremor cycle at which the stimulus is presented (145). In spite of variability, relatively high intercorrelations among different kinds of RT tests have been obtained (42), which suggests a common factor present in the various tests, and, according to Seashore, Buxton, and McCollom (130), such a factor can be analyzed. Further evidence supporting the notion of an RT factor comes from studies showing no relationship between RT and intelligence (42, 133), substitution (133), card sorting (133), discrimination reaction time (42, 130), pursuit rotor performance (42, 130), and little or no relationship to tapping tests (130). Slocombe and Brakeman (134) have also suggested that RT measures yield a group factor and, in addition, that this factor may be used to discriminate, at least among motormen, between good and poor accident risks.

SPECIAL FACTORS

Certain conditions, such as exposure to extreme climatic conditions, the effects of drugs, prolonged effort, starvation, deprivation of sleep for long periods, etc., are singled out here because of the unique interest which

they hold for many investigators. Some, if not all, of these could have been included in the previous section.

RT as a Function of Prolonged Readiness

Mackworth (102) in a series of studies of military personnel under conditions of prolonged vigilance found that the RT to the double-jumps of a clock hand increased sharply after one-half hour of watching. This increase could be prevented by the use of Benzedrine or by informing *S* of the results of the test.

In a related study Kennedy and Travis (87) measured the action potentials and the RT to combined light and sound stimuli under long monotonous conditions. Their results indicate that the frequency of action potentials decreases and the RT to irregularly presented stimulation becomes longer as the testing period is lengthened. After falling asleep, one stimulus presentation was sufficient to awaken *S*, the RT on this occasion being considerably longer, as might be expected. However, the RT approached its normal level rapidly with subsequent presentations.

RT as a Function of Certain Common Drugs

The RT has been used as the "I told you so" of all those who would restrict the use of the common drugs. In spite of this, the effects of these drugs on RT are not always clear when the conditions of the experiment are carefully considered. Some of the experimental problems are discussed by Hull (77) and more recently by Gray and Trowbridge (59) and by Miller (109). No real attempt will be made here to evaluate this type of effect but merely to point out a few representative papers.

With respect to coffee drinking, Hawk (68) and Gilliland and Nelson (58) claimed lengthened RT's. On the other hand, there are studies such as that of Thornton *et al.* (143) in which no effect was claimed. Cigarette smoking appears to decrease the variability of visual RT's according to both Hull (77) and Fay (43), but has no other reliable effect. Alcohol, on the other hand, both increases variability and lengthens visual (108, 150) and auditory (150) RT's. Benzedrine appears to have little or no effect on the auditory RT (143). Mackworth (102), however, did find that Benzedrine tended to offset decrements produced by prolonged vigilance. Aspirin has no effect on either visual or auditory RT's (31). A 30 per cent saturation of carbon monoxide is required to produce an increase in visual RT (49). Finally, morphine has the effect of first shortening and then lengthening RT except when taken in large doses, in which case only the latter effect occurs (101).

RT as a Function of Temperature

A number of studies (e.g., 75, 110, 157) have been done to determine the effects of climatic stress, temperature in particular, on the RT. The general result of these studies is that ambient temperatures between a range of -50°F. and 117°F. have little or no effect on either RT or more complex reaction times. This conclusion was reached by Forlano, Barmack, and Coakley (50) after a careful review of the effects of ambient and body temperatures on RT. Most of the studies available for evaluation, however, are distinguished by the degree with which several main variables are confounded in one experiment, and consequently are given to difficulty of interpretation. Such a conclusion,

therefore, should not be accepted as firmly established.

RT's have also had a little attention with regard to skin temperatures. Craik and Macpherson (29) report that cooling of the hand with which the response is made may increase RT by 10-15 per cent. This conclusion is not reasonable on the basis of their study since an increase of this much turns out to be an increase of 0.02 to 0.06 sec., a change which has little significance in terms of the likely error of measurement and the size of the sample (two Ss).

A few RT studies have been done with body temperature and/or time of day as the independent variable. In general, these studies (89, 90, 104) suggest that RT exhibits a slight diurnal variation, but with large individual differences. The data may also be interpreted to indicate that RT is a function of body temperature and is only spuriously correlated with time of day (50).

Effects of Sleep Conditions on RT

Most of the studies in which the stimulus for the RT has been presented during sleep have been related to the problem of determining the onset or the depth of sleep. Since the detection of both of these conditions still lacks other independent criteria, studies of this sort have little validity with respect to generalizations about the effect of sleeping on the RT. A discussion of this problem may be found in Kleitman (88).

Mullin and Kleitman (113) used a criterion of sleep onset which was independent of the RT and then investigated the change in a verbal, auditory RT for periods of time following onset. In this experiment, onset of sleep was considered to be established when S released a piece of paper which he held in his hand.

Using both normal and feeble-minded adults and normal children as Ss, Mullin and Kleitman found a slow increase in the verbal RT for the first 25 min. after onset, followed by a stable period of long RT's, and then by a third period of rapid shortening of the RT. This cycle, which was sigmoid in nature, was completed during the first hour after onset and did not reappear during the rest of the night.

Considerably more work has been done concerning the effect of lack of sleep on RT. Although most of the studies were concerned primarily with more complex reaction times, enough have been done with the RT to allow some evaluation.

As long ago as 1896 Patrick and Gilbert (116) tested the auditory and visual RT's of three Ss deprived of sleep for 90 hours and reported a marked slowing of speed. Later, Robinson and Herrman (126), in a more rigorous experiment, found that experimental insomnia produced no consistent effect on RT. With the exception of one or two studies reviewed by Kleitman (88), this latter result has marked the literature ever since, not only with respect to RT measures but practically all measures taken. Part of the difficulty might be blamed on the small samples usually used. On the other hand, large samples are administratively extremely difficult to achieve in this type of investigation. Lee and Kleitman (93), for example, tested the RT of only one S following 114 hours of sleep deprivation. This one S showed no effect. Cooperman, Mullin, and Kleitman (28) report no change in auditory RT for six Ss after 60 hours of privation, as did Tyler (149) after 24-114 hours of sleep loss.

Edwards (36) reported a study involving relatively large samples. He

compared the auditory RT of 17 Ss deprived of sleep for 100 hours with that of ten control Ss. His data support the studies already cited since he found no significant differences between the groups. It seems, therefore, that although extremely small samples may contribute to the variability of results, they cannot be considered the major reason for the failure of more recent studies to confirm the results of Patrick and Gilbert.

Most writers, when discussing the general failure to show that sleep loss influences RT, invoke "compensation" on the part of S as an explanatory concept. What is usually meant is that S achieves the same performance under stress as he does when not under stress by expending greater effort. Granting the looseness of this argument, the possibility of testing it appears available. If one is willing to accept greater muscular tension during the foreperiod and during the RT interval as a measure of greater effort, it should be relatively easy to determine by measuring tension during these intervals if compensation is really a useful concept.

Other Special Effects

One experiment was found in which the effect on the RT of radial acceleration was studied. Canfield, Comrey, and Wilson (16) investigated the effects of positive acceleration forces of 1, 3, and 5 g on the RT's to light and also to sound. Both kinds of RT were found to lengthen significantly with increased acceleration. Moreover, the slowing of reaction observed was not due to decrements in sensory functioning as the Ss were neither "blacked out" nor "greyed out" and, in addition, the Ss did not notice any change in sensory functioning.

In a different type of study, Tuttle *et al.* (148) reported that omission

of breakfast increased the visual RT of their five Ss. In a more careful nutrition study involving restriction of vitamin B, Brožek *et al.* (12) found that partial restriction of vitamin B intake for 23 days had no effect on the auditory RT of their eight Ss. Only prolonged and severe deprivation produced significant decrements.

McFarland (97, 98, 99, 100) has shown that RT is affected by altitude (anoxia), but not significantly before at least 20,000 ft. Variability of RT, however, begins to increase at 2000-3000 ft. before the average RT shows a decrement. McFarland (97) has also reported that people who live on mountains at extreme altitudes have longer and more variable RT's than people living at sea level in the same region.

SUMMARY AND CONCLUSIONS

Factors other than those discussed would appear to be important, but have not been studied in relation to the RT. For example, it might be expected that RT would decrease as a function of speed of travel or stimulation across the retina or the skin. The condition of the receptor and of the responding member are still among the conditions which have received insufficient attention. In spite of a long history of research, the effect of the sensory system is still an open question.

Many of our notions about the RT depend upon what are now classical studies, studies which did not have the advantage of modern statistical procedures. The question of mono-versus bisensory stimulation falls in this class. Research is also required to ascertain the effect of stimulus duration and the interaction between duration and intensity. Related to this is the problem of area or number of receptors stimulated, and

this too is in need of further experimental investigation.

On the positive side, intensity functions are relatively well established. The last 20 years have yielded a considerable amount of information about the effects of age and sex. They have also produced more definitive information about the role of the responding member, and about the factors involved in the foreperiod, although a great deal more research must yet be performed before the latter is really well understood.

The incentive to study the effects of sleep loss, drugs, temperature, altitude, acceleration, vigilance, etc. to a large extent has come from applied areas. Few, if any, safe generalizations are yet available. However, these topics will undoubtedly continue to receive a great deal of attention since many of them constitute important military problems.

All things considered, the last two decades have been quite productive, not of mathematically expressed empirical laws, but of useful nonmathematical generalizations and of clearer definition of the problems than was available before. The present status of the RT study may best be evaluated in terms of the generalizations which it has yielded. In the opinion of the writer the following generalizations appear to have been reasonably well established.

1. There is a positive correlation between the visual and the auditory RT.

2. Simultaneous stimulation of more than one sense modality produces faster RT's than stimulation of just one. On the other hand, successive stimulation of different senses produces slower RT's than stimulation of a single sensory channel.

3. For visual and thermal RT's the greater the extent of the stimulus in

space, i.e. the greater the number of receptors stimulated, the faster the speed of reaction up to some limit.

4. Under daylight or illuminated conditions the visual RT becomes longer the greater the distance of stimulation from the fovea.

5. In the case of each receptor system, RT is a negatively accelerated decreasing function of intensity up to some maximum intensity value after which RT either becomes suddenly lengthened, the function at this point being discontinuous, or asymptotic to a physiological limit.

6. RT is a slowly falling growth function of chronological age until about 30 years after which it is a

slowly rising function.

7. In general the RT of the human male is faster than that of the female.

8. The optimum foreperiod of RT may be thought of as lying in a range between approximately 1.5 and 8.0 sec. Its position in this range is determined by a large number of factors including the duration and intensity of the warning signal and of the stimulus, and the amount, locus, and time of production of muscular tension.

9. RT is not related to the length, direction, or speed of movement of the responding member.

10. Under vigilance conditions, the longer the period during which S must respond, the longer the RT.

REFERENCES

- ABEL, THEODORA M. The influence of visual and auditory patterns on tactual recognition. *Amer. J. Psychol.*, 1934, **46**, 443-447.
- ARNOLD, D. C., & TINKER, M. A. The fixation pause of the eyes. *J. exp. Psychol.*, 1939, **25**, 271-280.
- ATWELL, W. O., & ELBEL, E. R. Reaction time of male high school students in 14-17 year age groups. *Res. Quart. Amer. Ass. Hlth.*, 1948, **19**, 22-29.
- BAXTER, B. A. A study of reaction time using factorial design. *J. exp. Psychol.*, 1942, **31**, 430-437.
- BAXTER, B., & TRAVIS, R. C. The reaction time to vestibular stimuli. *J. exp. Psychol.*, 1938, **22**, 277-282.
- BELLIS, C. J. Reaction time and chronological age. *Proc. Soc. exp. Biol. Med.*, 1933, **30**, 801-803.
- BERGER, G. O. Ueber den Einfluss der Reizstärke auf die Dauer ein facher psychischer Vorgänge mit besonderer Rücksicht auf Lichtreize. *Phil. Stud.*, 1886, **3**, 38-93.
- BILLS, A. G. Studying motor functions and efficiency. In T. G. Andrews, (Ed.), *Methods of psychology*. New York: Wiley, 1947.
- BLISS, C. B. Investigations in reaction time and attention. *Stud. Yale psychol. Lab.*, 1893, **1**, 1-55.
- BORING, E. G. *Physical dimensions of consciousness*. New York: Century, 1933.
- BREITWIESER, J. V. Attention and movement in reaction time. *Arch. Psychol.*, 1911, No. 4.
- BROWER, D., & SANDS, H. Relations between reaction time and personal adjustment as measured by the Bell Adjustment Inventory. *J. gen. Psychol.*, 1948, **38**, 229-233.
- BROWN, J. S., & SLATER-HAMMEL, H. T. Discrete movements in the horizontal plane as a function of their length and direction. *J. exp. Psychol.*, 1949, **39**, 84-95.
- BROŽEK, J., GUETZKOW, H., & KEYS, A. A. A study of the personality of normal young men maintained on restricted intakes of vitamin B complex. *Psychosom. Med.*, 1946, **8**, 98-109.
- CANFIELD, A. A., JR. The influence of positive *g* on reaction time. *Amer. Psychologist*, 1950, **5**, 362. (Abstract)
- CANFIELD, A. A., COMREV, A. L., & WILSON, R. C. A study of reaction time to light and sound as related to increased positive radial acceleration. *J. Aviat. Med.*, 1949, **20**, 350-355.
- CARLSON, W. R., DRIVER, R. C., & PRESTON, M. G. Judgment times for the method of constant stimuli. *J. exp. Psychol.*, 1934, **17**, 113-118.
- CASSEL, E. E., & DALLENBACH, K. M. The effect of auditory distraction upon the sensory reaction. *Amer. J. Psychol.*, 1918, **29**, 129-143.
- CATTELL, J. MCK. The influence of the

- intensity of the stimulus on the length of the reaction time. *Brain*, 1886, 8, 512-515.
20. CATTELL, J. MCK. On reaction time and velocity of the nerve impulse. *Nat. Acad. sci. Mem.*, 1893, 7, 410.
 21. CHERNIKOFF, R., & BROGDEN, W. J. The effect of response termination of the stimulus upon reaction time. *J. comp. physiol. Psychol.*, 1949, 42, 357-364.
 22. CHERNIKOFF, R., GREGG, L. W., & BROGDEN, W. J. The effect of fixed duration stimulus magnitude upon reaction time to a response terminated stimulus. *J. comp. physiol. Psychol.*, 1950, 43, 123-128.
 23. CHERNIKOFF, R., & TAYLOR, F. V. Reaction time to kinesthetic stimulation resulting from sudden arm displacement. *J. exp. Psychol.*, 1952, 43, 1-8.
 24. CHOCHOLLE, R. Etude de la psychophysiologie de l'audition par la méthode des temps de réaction. *Année Psychol.*, 1948, 45-56, 90-131.
 25. CHOCHOLLE, R. Quelques remarques sur les variations et la variabilité des temps de réaction auditifs. *J. Psychol. norm. path.*, 1948, 41, 345-358.
 26. COBB, P. W. Some experiments on speed of vision. *Trans. illum. engng Soc.*, 1924, 19, 150-175.
 27. COERMANN, R. Untersuchungen über die Einwirkung von Schwingungen auf den menschlichen Organismus. *Industr. Psychotech.*, 1939, 16, 169-206.
 28. COOPERMAN, N. R., MULLIN, F. J., & KLEITMAN, N. Studies on the physiology of sleep. XI. Further observations on the effects of prolonged sleeplessness. *Amer. J. Physiol.*, 1934, 107, 589-593.
 29. CRAIK, K. J. W., & MACPHERSON, S. J. Effects of cold upon hand movement and reaction time. Report Comm. Armoured Vehicles, MPRCBPC 43/196, March 13, 1943.
 30. DANIEL, R. S. Some observations on Meyer's study of reaction time and muscle tension. *J. exp. Psychol.*, 1949, 39, 896-898.
 31. DAVIS, R. C. The effects of analgesic dosage of aspirin (acetyl salicylic acid) on some mental and motor performances. *J. appl. Psychol.*, 1936, 20, 481-487.
 32. DAVIS, R. C. Motor components of responses to auditory stimuli: the relation of stimulus intensity and instructions to respond. *Amer. Psychologist*, 1947, 2, 308. (Abstract)
 33. DAVIS, R. C. Motor effects of strong auditory stimuli. *J. exp. Psychol.*, 1948, 38, 257-275.
 34. DAVIS, R. C. Motor responses to auditory stimuli above and below threshold. *J. exp. Psychol.*, 1950, 40, 107-120.
 35. EDES, B., & DALLENBACH, K. M. The adaptation of pain aroused by cold. *Amer. J. Psychol.*, 1936, 48, 307-315.
 36. EDWARDS, A. S. Effects of the loss of one hundred hours of sleep. *Amer. J. Psychol.*, 1941, 54, 80-91.
 37. EDWARDS, W. Recent research on pain perception. *Psychol. Bull.*, 1950, 47, 449-474.
 38. ELLIOT, F. R., & LOUITTIT, C. M. Auto braking reaction times to visual vs. auditory warning signals. *Proc. Ind. Acad. Sci.*, 1948, 47, 220-225.
 39. ELLSON, D. G., & HILL, H. The interaction of responses to step function stimuli. I. Opposed steps of constant amplitude. *USAF, Aero Med. Lab., Wright-Patterson AFB, MCREXD* 694-28, 1948.
 40. ESCHER-DESRIVIERES, J. Variations des temps de réactions psychomotrices visuelles en fonction de l'éclairement en lumière blanche et colorée. *C. R. Acad. Sci., Paris*, 1939, 208, 1751-1753.
 41. FARMER, E., & CHAMBERS, E. G. A psychological study of individual differences in accident rates. *Med. Res. Coun., Industr. Hlth Res. Bd.*, Report No. 38, 1926, London, Eng.
 42. FARNSWORTH, P. R., SEASHORE, R. H., & TINKER, M. A. Speed in simple and serial action as related to performance in certain "intelligence" tests. *J. genet. Psychol.*, 1927, 34, 537-551.
 43. FAY, P. J. The effect of cigarette smoking on simple and choice reaction time to colored lights. *J. exp. Psychol.*, 1936, 19, 592-603.
 44. FÉRÉ, F. L. L'énergie et la vitesse des mouvements volontaires. *Rev. Phil.*, 1889, 28, 36-69.
 45. FERREE, C. E., & RAND, G. Intensity of light and speed of vision studied with special reference to industrial situations. Part I. *Trans. illum. engng Soc.*, 1927, 22, 79-110.
 46. FINAN, J. L., FINAN, S. C., & HARTSON, L. D. A review of representative tests used for the quantitative measure.

- ments of behavior-decrement under conditions related to aircraft flight. Dayton, O.: U. S. Air Material Command, Wright-Patterson Air Force Base, 1949. iv, 230 p. (USAF Tech. Rep. No. 5830.)
47. FINCH, G. Review of muscle activity and action potentials as they are related to movement. (AAF AMC Aero Med. Lab. Memo Rep. TSEAA 694-2E, 1947; Publ. Bd. No. M 81423.) Washington, D. C.: U. S. Dep. Commerce, 1947.
 48. FORBES, G. The effect of certain variables on visual and auditory reaction times. *J. exp. Psychol.*, 1945, **35**, 153-162.
 49. FORBES, W. H., DILL, D. B., DE SILVA, H., & VANDEVENTER, F. M. The influence of moderate carbon monoxide poisoning upon the ability to drive automobiles. *J. indust. Hyg. Toxicol.*, 1937, **19**, 598-608.
 50. FORLANO, G., BARMACK, J. E., & COAKLEY, J. D. The effect of ambient and body temperatures upon reaction time. *Special Devices Center, Report No. 151-1-13*, 1948.
 51. FRANKLIN, J. C., & BROZEK, J. The relation between distribution of practice and learning efficiency in psychomotor performance. *J. exp. Psychol.*, 1947, **37**, 16-24.
 52. FREEMAN, G. L. The optimal locus of anticipatory tensions in muscular work. *J. exp. Psychol.*, 1937, **21**, 554-564.
 53. FREEMAN, G. L. The optimal muscular tensions for various performances. *Amer. J. Psychol.*, 1938, **51**, 146-150.
 54. FREEMAN, G. L., & KENDALL, W. E. The effect upon reaction time of muscular tension induced at various preparatory conditions. *J. exp. Psychol.*, 1940, **27**, 136-148.
 55. FROEBERG, S. The relation between the magnitude of stimulus and the time of reaction. *Arch. Psychol.*, 1907, **16**, No. 8, 1-38.
 56. FROELICH, F. W. *Die Empfindungszeit*. (Ed. 1) Jena: Fischer, 1929.
 57. GALIFRET, Y., & PIÉRON, H. Vitesse de réaction et intensité de sensation. Données expérimentales sur le problème d'une courbe sigmoïde des vitesses. *Année Psychol.*, 1951, **51**, 1-16.
 58. GILLILAND, A. R., & NELSON, D. The effects of coffee on certain mental and physiological functions. *J. gen. Psychol.*, 1939, **21**, 339-348.
 59. GRAY, M. G., & TROWBRIDGE, E. L. Methods for investigating the effects of drugs on psychological function. *Psychol. Rev.*, 1942, **5**, 127-148.
 60. GREENSHIELDS, B. D. Reaction time in automobile driving. *J. appl. Psychol.*, 1936, **20**, 353-358.
 61. GREGG, E. C., JR. Physical basis of pain threshold measurements in man. *J. appl. Physiol.*, 1951, **4**, 351-363.
 62. GREGG, L. W., & BROGDEN, W. J. The relation between duration and reaction time difference to fixed duration and response terminated stimuli. *J. comp. physiol. Psychol.*, 1950, **43**, 329-337.
 63. GREGG, L. W., & BROGDEN, W. J. The relation between reaction time and the duration of the auditory stimulus. *J. comp. physiol. Psychol.*, 1950, **43**, 389-395.
 64. GUILFORD, J. P., & EWART, E. Reaction time during distraction as an indication of attention-value. *Amer. J. Psychol.*, 1940, **53**, 554-563.
 65. HAMEL, I. A. A study and analysis of the conditioned reflex. *Psychol. Monogr.*, 1919, **27**, No. 1 (Whole No. 118).
 66. HATHAWAY, S. R. An action potential study of neuromuscular relations. *J. exp. Psychol.*, 1935, **11**, 285-298.
 67. HATHAWAY, S. R., & SISSON, E. D. The time relations of the events in quick voluntary movements. *Psychol. Bull.*, 1935, **32**, 721-722.
 68. HAWK, P. B. A study of the physiological and psychological reactions of the human organism to coffee drinking. *Amer. J. Physiol.*, 1929, **90**, 380-381.
 69. HENDERSON, R. L. Remote action potentials at the moment of response in a simple reaction-time situation. *J. exp. Psychol.*, 1952, **44**, 238-241.
 70. HENMON, V. A. C., & WELLS, F. L. Concerning individual differences in reaction times. *Psychol. Rev.*, 1914, **21**, 153-156.
 71. HENRY, F. M. Independence of reaction and movement times and equivalence of sensory motivators of faster response. *Res. Quart. Amer. Ass. Hlth*, 1952, **23**, 43-53.
 72. HICK, W. E. Reaction time for the amendment of a response. *Quart. J. exp. Psychol.*, 1949, **1**, 175-179.
 73. HILDEN, A. H. An action current study of the conditioned hand withdrawal. *Psychol. Monogr.*, 1937, **49**, No. 1 (Whole No. 217), 173-204.
 74. HOLMES, J. L. Reaction time to light as conditioned by wave-length and in-

- tensity. Unpublished doctor's dissertation, Columbia Univer., 1923.
75. HORVATH, S. M., & FREEDMAN, A. The influence of cold upon the efficiency of man. *J. Aviat. Med.*, 1947, **18**, 158-164.
 76. HOVLAND, C. I. The influence of adaptation illumination upon visual reaction time. *J. gen. Psychol.*, 1936, **14**, 346-359.
 77. HULL, C. L. The influence of tobacco smoking on mental and motor efficiency. *Psychol. Monogr.*, 1924, **33**, No. 3, (Whole No. 150), 1-160.
 78. HULL, C. L. Stimulus intensity dynamism (V) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67-76.
 79. JENKINS, T. N. Facilitation and inhibition. *Arch. Psychol.*, 1926, No. 86, 1-56.
 80. JOHANSON, A. M. The influence of incentive and punishment upon reaction-time. *Arch. Psychol.*, 1922, No. 54, 1-52.
 81. JOHNSON, H. M. The influence of the distribution of brightnesses over the visual field on the time required for discriminative responses to visual stimuli. *Psychobiol.*, 1918, **1**, 459-494.
 82. JOHNSON, H. M. Reaction time measurements. *Psychol. Bull.*, 1923, **20**, 562-589.
 83. JONES, B. F., FLINN, R. H., HAMMOND, E. C., *et al.* Fatigue and hours of service of Interstate Truck Drivers. *Pub. Hlth Bull.*, 1941, No. 265, Fed. Sec. Agency, U. S. Pub. Hlth. Ser., Wash., D. C.
 84. JONES, H. E. *Motor performance and growth*. Berkeley: Univer. of California Press, 1949.
 85. JUDD, C. H., MCALLESTER, C. H., & STEELE, W. M. Analysis of reaction movements. *Psychol. Monogr.*, 1904, **7**, No. 1, (Whole No. 29), 141-184.
 86. KENNEDY, J. L., & TRAVIS, R. C. Prediction and control of alertness. II. Continuous tracking. *J. comp. physiol. Psychol.*, 1947, **41**, 203-210.
 87. KENNEDY, J. L., & TRAVIS, R. C. Prediction of speed of performance by muscle action potentials. *Science*, 1947, **105**, 410-411.
 88. KLEITMAN, N. *Sleep and wakefulness*. Chicago: Univer. of Chicago Press, 1939.
 89. KLEITMAN, N., & JACKSON, O. P. Body temperature and performance under different routines. *J. appl. Physiol.*, 1950, **3**, 304-328.
 90. KLEITMAN, N., TITELBAUM, S., & FEIVESON, P. The effect of body temperature on reaction time. *Amer. J. Physiol.*, 1938, **121**, 495-501.
 91. LANDAHL, H. D. Contributions to the mathematical biophysics of the central nervous system. *Bull. math. Biophysics*, 1939, **1**, 95-118.
 92. LANIER, L. H. The interrelations of speed and reaction measurements. *J. exp. Psychol.*, 1934, **17**, 371-399.
 93. LEE, M. A. M., & KLEITMAN, N. Studies on the physiology of sleep. II. Attempts to demonstrate functional changes in the nervous system during experimental insomnia. *Amer. J. Physiol.*, 1923, **67**, 141-151.
 94. LEMMON, V. W., & GEISINGER, S. M. Reaction time to retinal stimulation under light and dark adaptation. *Amer. J. Psychol.*, 1936, **48**, 140-142.
 95. LIVINGSTON, W. A. Action potential measurements from the arm in the foreperiod of reaction time to visual stimuli. *Proc. Ind. Acad. Sci.*, 1946, **55**, 170. (Abstract)
 96. LUCKIESH, M. *Light, vision, and seeing*. New York: Van Nostrand, 1944.
 97. MCFARLAND, R. A. The psychological effects of oxygen deprivation (anoxemia) on human behavior. *Arch. Psychol.*, 1932, No. 145, 1-135.
 98. MCFARLAND, R. A. Psycho-physiological studies at high altitude in the Andes. I. The effect of rapid ascents by aeroplane and train. *J. comp. Psychol.*, 1937, **23**, 191-225.
 99. MCFARLAND, R. A. Psycho-physiological studies at high altitude in the Andes. II. Sensory and motor responses during acclimatization. *J. comp. Psychol.*, 1937, **23**, 227-258.
 100. MCFARLAND, R. A. Psycho-physiological studies at high altitude in the Andes. IV. Sensory and circulatory responses of the Andean residents at 17,500 ft. *J. comp. Psychol.*, 1937, **24**, 189-220.
 101. MACHT, D. I., & ISAACS, S. Action of some opium alkaloids on the psychological reaction time. *Psychobiol.*, 1917, **1**, 19-32.
 102. MACKWORTH, N. H. *Researches on the measurement of human performance*. London: His Majesty's Stationery Office, 1950. (Med. Res. Coun., Special Rep. Ser., No. 268.)
 103. MALLORY, E. B. The recognition of relatively simple sensory experiences. *Amer. J. Psychol.*, 1943, **46**, 120-131.

104. MARSH, H. D. The diurnal course of efficiency. *Arch. Phil. Psychol. Sci. Methods*, 1906, No. 7.
105. MEYER, H. D. Reaction time as related to tensions in muscles not essential in the reaction. *J. exp. Psychol.*, 1949, **39**, 96-113.
106. MEYER, H. D. Some remarks concerning Daniel's observations. *J. exp. Psychol.*, 1949, **39**, 898-900.
107. MILES, W. R. Correlation of reaction and coordination speed with age in adults. *Amer. J. Psychol.*, 1931, **43**, 377-391.
108. MILES, W. R. *Alcohol and human efficiency*. Washington, D. C.: Carnegie Inst., 1936.
109. MILLER, L. C. A critique of analgesic testing methods. *Ann. N. Y. Acad. Sci.*, 1948, **51**, 34-50.
110. MITCHELL, H. H., GLICKMAN, U., LAMBERT, E. H., KEETON, R. W., & FAHNESTOCK, M. K. The tolerance of man to the cold as affected by dietary modification: carbohydrate versus fat and the effect on the frequency of meals. *Amer. J. Physiol.*, 1946, **146**, 84-96.
111. MOORE, T. V. A study of reaction time and movement. *Psychol. Rev. Monogr. Suppl.*, 1904, **6**, No. 1 (Whole No. 24).
112. MOWRER, O. H. Preparatory set (expectancy)—some methods of measurement. *Psychol. Monogr.*, 1940, **52**, No. 2 (Whole No. 233).
113. MULLIN, F. J., & KLEITMAN, N. Variations in threshold of auditory stimuli necessary to awaken the sleeper. *Amer. J. Physiol.*, 1938, **123**, 477-481.
114. MÜNNICH, K. Die Reaktionsleistung in Abhängigkeit von der Körperlage. *Industr. Psychotech.*, 1940, **17**, 49-83. (See *Psychol. Abstr.*, 1947, No. 1649.)
115. MUSCIO, B. On the relation of fatigue and accuracy to speed and duration of work. *Med. Res. Coun., Industr. Hlth Res. Bd.*, 1922, Report No. 19-B, London, Eng.
116. PATRICK, G. T. W., & GILBERT, J. A. On the effects of loss of sleep. *Psychol. Rev.*, 1896, **3**, 469-483.
117. PATTLE, R. E., & WEDDELL, G. Observations on electrical stimulation of pain fibres in an exposed human sensory nerve. *J. Neurophysiol.*, 1948, **11**, 93-98.
118. PIÉRON, H. Nouvelles recherches sur l'analyse du temps de latence sensorielle et sur la loi qui relie le temps à l'intensité d'excitation. *Année Psychol.*, 1920, **22**, 58-142.
119. PIÉRON, H. Recherches expérimentales sur la marge de variation du temps de latence de la sensation lumineuse (par une méthode de masquage). *Année Psychol.*, 1926, **26**, 1-30.
120. PIÉRON, H. *The sensations: their functions, processes and mechanisms*. (Trans. by M. H. Pierenne & B. C. Abbott.) New Haven: Yale Univer. Press, 1952.
121. POFFENBERGER, A. T. Reaction time to retinal stimulation, with special reference to the time lost in conduction through nerve centers. *Arch. Psychol.*, 1912, No. 23, 1-73.
122. POSTMAN, L., & KAPLAN, H. L. Reaction time as a measure of retroactive inhibition. *J. exp. Psychol.*, 1947, **37**, 136-145.
123. POULTON, E. C. Perceptual anticipation and reaction time. *Quart. J. exp. Psychol.*, 1950, **2**, 99-112.
124. RASHEVSKY, N. *Advances and applications of mathematical biology*. Chicago: Univer. of Chicago Press, 1940.
125. ROBINSON, E. S. Work of the integrated organism. In C. Murchison (Ed.), *Handbook of general experimental psychology*. Worcester, Mass.: Clark Univer. Press, 1934.
126. ROBINSON, E. S., & HERMANN, S. O. Effects of loss of sleep. *J. exp. Psychol.*, 1932, **15**, 19-32.
127. ROOS, J. The latent period of skeletal muscle. *J. Physiol.*, 1932, **74**, 17-33.
128. SALTZMAN, I. J., & GARNER, W. R. Reaction time as a measure of span of attention. *J. Psychol.*, 1948, **25**, 227-241.
129. SEARLE, L. V., & TAYLOR, F. V. Studies of tracking behavior. I. Rate and time characteristics of simple corrective movements. *J. exp. Psychol.*, 1948, **38**, 615-631.
130. SEASHORE, R. H., BUXTON, C. E., & MCCOLLOM, I. N. Multiple-factor analysis of fine motor skills. *Amer. J. Psychol.*, 1940, **53**, 251-259.
131. SEASHORE, R. H., STARMAN, R., KENDALL, W. E., & HELMICK, J. S. Group factors in simple and discriminative reaction times. *J. exp. Psychol.*, 1941, **29**, 346-349.
132. SEASHORE, S. H., & SEASHORE, R. H. Individual differences in simple auditory reaction times of hands, feet, and jaws. *J. exp. Psychol.*, 1941, **29**, 342-345.
133. SISK, T. K. The interrelations of speed in simple and complex responses.

- Peabody Coll. Contr. Educ., 1926, No. 23.
134. SLOCOMBE, C. S., & BRAKEMAN, E. E. Psychological tests and accident proneness. *Brit. J. Psychol.*, 1930, **21**, 29-38.
 135. SMITH, K. U. The functions of the inter-cortical neurones in sensorimotor co-ordination and thinking in man. *Science*, 1947, **105**, 234-235.
 136. SMITH, W. M. Sensitivity to apparent movement in depth as a function of stimulus dimensionality. *J. exp. Psychol.*, 1952, **43**, 149-155.
 137. STEINMAN, A. R. Reaction time to change compared with other psychophysical methods. *Arch. Psychol.*, New York, 1944, No. 292, 34-60.
 138. STEINMAN, A., & VENIAR, S. Simple reaction time to change as a substitute for the disjunctive reaction. *J. exp. Psychol.*, 1944, **34**, 152-158.
 139. STONE, L. J., & DALLENBACH, K. M. Adaptation to the pain of radiant heat. *Amer. J. Psychol.*, 1934, **46**, 229-242.
 140. STRUGHOLD, H. The human time factor in flight. The latent period of optical perception and its significance in high speed flying. *J. Aviat. Med.*, 1949, **20**, 300-307.
 141. STRUGHOLD, H. The human time factor in flight: II. Chains of latencies in vision. *J. Aviat. Med.*, 1951, **22**, 100-108.
 142. TELFORD, C. W. The refractory phase of voluntary and associative responses. *J. exp. Psychol.*, 1931, **14**, 1-36.
 143. THORNTON, G. R., HOLCK, H. G. O., & SMITH, E. L. The effect of benzedrine and caffeine upon performance in certain psychomotor tasks. *J. abnorm. soc. Psychol.*, 1939, **34**, 96-113.
 144. THURSTONE, L. L. Psychophysical methods. In T. G. Andrews (Ed.), *Methods of psychology*, New York: Wiley, 1947.
 145. TIFFIN, J., & WESTHAFFER, F. L. The relation between reaction time and temporal location of the stimulus on tremor cycle. *J. exp. Psychol.*, 1940, **27**, 318-324.
 146. TODD, J. W. Reaction to multiple stimuli. *Arch. Psychol.*, 1912, No. 25, 1-65.
 147. TRAVIS, R. C., & KENNEDY, J. L. Prediction and automatic control of alertness. I. Control of lookout alertness. *J. comp. physiol. Psychol.*, 1947, **40**, 457-461.
 148. TUTTLE, W. W., WILSON, M., & DAUM, K. Effect of altered breakfast habits on physiologic response. *J. appl. Physiol.*, 1949, **1**, 545-559.
 149. TYLER, D. B. The effect of amphetamine sulfate and some barbiturates on the fatigue produced by prolonged wakefulness. *Amer. J. Physiol.*, 1947, **150**, 253-262.
 150. VARÉ, P. Influence de l'alcool sur les réactions psychométriques. *C. R. Soc. Biol.*, 1932, **11**, 70-72.
 151. VERBILLE, E. The effect of emotional and motivational sets on the perception of incomplete pictures. *J. genet. Psychol.*, 1946, **69**, 133-145.
 152. VINCE, M. A. Corrective movements in a pursuit task. *Quart. J. exp. Psychol.*, 1948, **1**, 85-103.
 153. WELLS, F. L., KELLEY, C. M., & MURPHY, G. Comparative simple reactions to light and sound. *J. exp. Psychol.*, 1921, **4**, 57-62.
 154. WELLS, F. L., KELLEY, C. M., & MURPHY, G. On attention and simple reaction. *J. exp. Psychol.*, 1921, **4**, 391-398.
 155. WELLS, G. R. The influence of stimulus duration on reaction time. *Psychol. Monogr.*, 1913, **15**, No. 5 (Whole No. 66).
 156. WERTHEIMER, M. A single-trial technique for measuring the threshold of pain by thermal radiation. *Amer. J. Psychol.*, 1952, **65**, 297-298.
 157. WILLIAMS, C. C., & KITCHING, J. A. The effects of cold on human performance. I. Reaction time. 1942, *Misc. Canad. Aviat. Rep.*, No. 81-A.
 158. WILLIAMS, R. D. Experimental analysis of forms of reaction movement. *Psychol. Monogr.*, 1914, **17**, No. 4 (Whole No. 75), 55-155.
 159. WOODROW, H. The measurement of attention. *Psychol. Monogr.*, 1914, **17**, No. 5 (Whole No. 76).
 160. WOODROW, H. Reactions to the cessation of stimuli and their nervous mechanism. *Psychol. Rev.*, 1915, **22**, 423-452.
 161. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.
 162. WRIGHT, G. W. The latency of sensations of warmth due to radiation. *J. Physiol.*, 1951, **112**, 344-358.
 163. WUNDT, W. *Grundzüge der physiologischen Psychologie*. (5th Ed.) Leipzig: Engelmann, 1903.

Received June 5, 1953.

REPRESENTATIVE vs. SYSTEMATIC DESIGN IN CLINICAL PSYCHOLOGY

KENNETH R. HAMMOND

University of Colorado

The purpose of this article is to illustrate how the application of traditional, *systematic*, rather than *representative*, experimental design to problems in clinical psychology results in unjustified conclusions.¹ Five examples, all drawn from attempts to discover the effect of the examiner on the subjects' responses, are presented. In each case it will be shown that justified conclusions regarding the problems being investigated could have been drawn if representative rather than systematic design had been used. The presentation is intended to be illustrative rather than exhaustive.

Consider first the simplest form of systematic design—the classical one-variable design. Both Brunswik (3, p. 8) and Fisher (11, p. 88) point out that this design is inherited from classical physics and both agree that revision of this procedure is necessary. Fisher, for example, remarks:

In expositions of the scientific use of experimentation it is frequent to find an exces-

¹ The two types of design are juxtaposed here in conformity with a distinction established by Brunswik (3). In the sense in which he uses the terms, "representative design" refers to the transfer of the principles of sampling statistics from the subjects of a psychological investigation to the objects or situations which constitute the stimuli in the investigation. The arbitrary orderliness with which these external (independent) variables are customarily handled is summarized by Brunswik under the opposite heading of "systematic design," and Fisher's factorial design (11) is presented as a relatively recent and relatively complex example. Fisher has also used the term "systematic" in a similar, albeit somewhat casual, manner (12).

sive stress laid on the importance of varying the essential conditions only one at a time. . . . This ideal doctrine seems to be more nearly related to expositions of elementary physical theory than to laboratory practice in any branch of research. In experiments merely designed to illustrate or demonstrate simple laws, connecting cause and effect, the relationships of which with the laws relating to other causes are already known, it provides a means by which the student may apprehend the relationship, with which he is to familiarize himself, in as simple a manner as possible. By contrast, in the state of knowledge or ignorance in which genuine research, intended to advance knowledge, has to be carried on, this simple formula is not very helpful (11, p. 88).

Brunswik and Fisher agree concerning the disadvantages of classical experimental design, but they differ as to the remedy. Their differences lie not so much in the general nature of the reform needed as in the extent of the reform. Thus, although Fisher's dissatisfaction led him to develop the multivariate analysis of variance and related techniques, Brunswik's efforts have resulted in a more thorough and more radical revision of experimental methodology. For, although Fisher urged multivariation of conditions in the experiment (wherein results are obtained), thereby approaching *one* of the conditions permitting the application of results, Brunswik urges that in order to eliminate the traditional artificiality in the choice and manner of variation of the experimental variables, the conditions of the experiment *represent statistically* the universe of situations toward which one wishes to generalize.

In suggesting that the logic of sampling theory (which psychologists have long applied to populations of subjects) be applied to stimulus situations, Brunswik brings experimental methodology in line with the modern statistical approach to the problem of inductive inference. For example, "it is clear that the statistical significance of a result may be investigated in both directions" (3, p. 36). Situational (or "ecological") generality of results, therefore, is as much a challenge to statistical scrutiny as the subject-populational generality of results. As Brunswik points out, not only should we be concerned with the number of stimulus variables included in the experiment, but we must also consider the *manner of variation of the stimulus variables and their covariation among one another*, a problem which Fisher, in his concern with the mathematical problems involved in increasing the number of variables, did not consider as fully as psychologists might wish.

The above is an extremely cursory description of the fundamental notion of representative design. The examples from clinical psychology presented later, however, will further clarify its meaning. For a complete exposition of representative design, the reader is referred to Brunswik's monograph (3). From the field of academic psychology—notably from the perceptual constancies, depth perception, gestalt problems and illusions, and social perception—the reader may find concrete examples of representatively designed experiments and ecological surveys in (2; 3, pp. 24–38, 41–52; 6; 9). The general basis for representative design as given by the probabilistic nature of behavioral adjustment is presented by Brunswik in (1). Further discussion of representative design and of

its place in the historical development of psychology may be found in (5).

An analogy between the discovery, on the part of relativity physics, of the confinement of traditional physical laws to a limited universe of conditions, on the one hand, and the considerations in psychology that have led to the establishment of representative design, on the other, has been pointed out by Hammond (16). In a note to this analogy, Brunswik (4) has endorsed the writer's interpretation that in spite of the apparent stress on "theory," in the first case, and on "design," in the second, the major problem in both cases is the same, that is, "generalization."

Another paper by Hammond (15) constitutes an application of the principles of representative design to certain generalization problems frequently mismanaged in current social and clinical psychology. This paper endeavored to criticize a concrete example of a research project, an interview study by Robinson and Rohde (21). In this study the subjects had been properly sampled but the sampling of the interviewers had been ignored; yet the conclusions concerning the "significance" of the results were extended to both the subject and the interviewer populations. The point made by the present writer in criticism of this report was that the interviewers should have been treated as "objects," that is, as parts of the "situations," and thus under the rules of representative design should have been sampled in the same manner as were the subjects proper; or else the generalizations should have been limited to the subject population.²

² In pointing out the wide discrepancies between the large numbers of "judges" (responders or subjects proper) and the small numbers of "social objects" (subjects func-

In the present paper we wish to extend our criticism of research reported in the literature to the inappropriate application of systematic design to the type of experiment which seeks to determine the effect of the examiner on the subjects' responses.

ONE-FACTOR SYSTEMATIC DESIGN

In the examples discussed below, the attempt is made to vary one factor, such as the race or sex of the examiner, in order to find the effect of the stimulus variable in question on the subjects' responses.

Effect of race. Reiss, Schwartz, and Cottingham (20) were concerned mainly with verifying the utility of the (Thompson) Negro version of the TAT, but were also concerned with discovering the effect of the race of the examiner on the responses of the subjects to the Thompson TAT. "Fifteen Negro and 15 white students were tested by a Negro administrator and 15 Negro and 15 white students were tested by a white administrator" (20, p. 704). Thus, the independent variable of race is constituted by *one* representative from each race. In light of that fact, consider the following conclusion drawn by the authors: "Northern white Ss produce longer stories than do Northern Negroes when the stimulus material offers Negro figures *regardless of the color of the examiner*" (italics ours) (20, p. 708).

Note that this conclusion follows from the comparison of the stories of white and Negro subjects to *one* white

tioning as stimuli) typically found in the literature on social perception from photographs, Brunswik (3, p. 38 and Table 2) has suggested the use of the lower case letter *n* for the size of the responder sample and of the capital letter *N* for the size of the ecological or object sample. This system has been used in the present paper also.

and *one* Negro examiner. It is apparent that while there is a potentially adequate sample of $n=60$ subjects, the size of the object (or ecological) sample is thus a bare $N=2$ (for the use of n vs. N see footnote 2). If, in fact, "it is clear that the statistical significance of a result may be investigated in both directions" (3, p. 36), then we may legitimately ask how it is possible to generalize from results obtained with one representative of each race as a stimulus, any more than if the experiment included only one member from each race as a subject? Generalization from sample to population is necessary on both sides of this experiment and the statistical rules which permit generalization hold under both conditions. For Reiss *et al.* to compare the results obtained under the conditions described above with Thompson's results is just as meaningless as if the two experiments were each carried out with a single subject. Clearly, an invalid generalization is drawn in the above experiment. Only a representative type of experimental design in which adequate consideration of the situation to which the experimenter wishes to generalize is given would permit valid generalization concerning the effect of the stimuli.

Effect of sex. Curtis and Wolf report an experiment in which the problem is stated as follows: "To study the influence of the sex of the examiner on the production of sex responses on the Rorschach" (8, p. 345). The subject-population sample consisted of 586 Rorschach records. The independent variable, sex of the examiner, was constituted by three female and seven male examiners. The conclusions are: "There is a significant difference between our male and female examiners on the number of records with sex responses."

This experiment suffers from the same lack of attention to the establishment of an independent variable as the experiment discussed earlier. Note the elaborate concern of the authors to establish subject-population generality ($n=586$). Contrast this sampling procedure with that on the stimulus side (male $N=7$, female $N=3$). Yet it is the latter sample which constitutes the independent variable and which therefore leads to the conclusion that the sex of the examiner influences the response of the subject. This type of imbalance (favoring n over N) is characteristic of the indiscriminate application of systematic design. Again, there is a clearly invalid generalization on the stimulus side of the experiment which only representative design can remedy.

An approximation to representative design. An example of an "effect of the examiner" experiment which approximates representative design may be found in Gibby (14). Twelve examiners each tested 20 patients under one set of conditions, and in the second set of conditions "one hundred thirty-five subjects were used, each randomly assigned to one of nine examiners" (14, p. 450). Although this experiment also shows marked imbalance (subject n 's equaling 240 and 135, and object N 's equaling 12 and 9), a better approximation is obtained here than in any other experiment of this sort seen by the writer. Gibby's experiment is cited mainly to illustrate that representative design does not present insuperable difficulties.

One further point. If the experimenter wishes to limit his conclusions to the particular conditions of the experiment (as is frequently the case in applied problems), we have no criticism. If, for example, Reiss *et al.*

wish to point out that *this* examiner produces (or does not produce) different responses than *that* examiner, we have no criticism. Generalizations such as those made in the above experiments, however, demand randomization—and randomization is what is lacking.

In summary, unwarranted conclusions concerning the effect of the examiner occur as a result of failure to establish the crucial independent variable. It is important to note that failure to establish the independent variable occurs as a result of applying the logic of statistical inference to one side of the experiment only.

MULTIFACTOR SYSTEMATIC DESIGN

This section is concerned with multifactor systematic design (factorial design) methods and representative design. We begin with an example from the same field of research as in the previous section.

Sex and method of administration. Garfield, Blek, and Melker (13) were concerned with ascertaining the effect of the sex of the examiner and two methods of administration (complete session and interrupted session) upon TAT stories. Two male and two female examiners constituted the sex variable and one of each was assigned to the different methods. The experimenters also wished to discover whether differential effects would be obtained with male ($n=54$) and female ($n=56$) subjects. Although the experimenters do not treat their results completely in terms of a factorial design, most readers will recognize that the conditions of the experiment allow use of a $2 \times 2 \times 2$ design as in Table 1. For purposes of illustration we will assume that Garfield *et al.* did set their experiment up in this fashion. This assumption will in no way invalidate our analysis.

TABLE 1
EXAMPLE OF FACTORIAL DESIGN

Split Session				Complete Session			
Male <i>E</i>		Female <i>E</i>		Male <i>E</i>		Female <i>E</i>	
Male <i>Ss</i>	Fe <i>Ss</i>	Male <i>Ss</i>	Fe <i>Ss</i>	Male <i>Ss</i>	Fe <i>Ss</i>	Male <i>Ss</i>	Fe <i>Ss</i>

A useful, although uncustomary, step in analyzing the independent variables of experiments designed in this fashion is to separate the physical condition variables from the "person condition" variables. If, for example, the independent variables include 2 methods, 2 types of examiners, and 2 types of subjects, as does the Garfield experiment, it is important to note that we have a set of physical conditions (methods) and two sets of "person" conditions. In the experiment under discussion, the "person" conditions can be further broken down into subjects and "objects," i.e., examiners.

Person conditions (subjects). The customary principles of subject-population sampling apply here and need no discussion.

Person condition (objects). Again it is easy to see that the same restrictions concerning generalization apply here as in subject sampling. Failure to observe these restrictions results in Garfield's unjustified generalization that sex of the examiner does not influence the subjects' response. Factorial design per se does not remove these restrictions. As Edwards says in discussing a similar hypothetical factorial design experiment wherein instructors and methods constitute the independent variables, "let us suppose that we have selected the instructors to represent particular types or personalities or abilities. The three used in the experiment are definitely not a random sample from

any defined population" (10, p. 249). In other words, generalization to other instructors is prohibited; conclusions are confined to the *particular* instructors involved in the experiment.

Here we would like to point out that in "effect of the examiner" experiments it ordinarily will be more important to achieve preciseness with regard to the independent variable than the dependent variable. That is, we will be primarily concerned with the problem of whether or not certain examiner variables make a difference—to whom (that is, to which subject-population) they make a difference will ordinarily be of less concern. Therefore sampling of objects (examiners) will require a larger sample, more carefully considered in terms of representativeness, than will sampling of subjects.

Physical conditions. Here the issue of sampling and generalization becomes somewhat obscure. It is not easy to conceive of methods, or tests, or other physical conditions being selected by means of random sampling procedures. Yet this issue of random selection of conditions has a very important and practical bearing on factorial methods, for it can be reduced to the question of what one considers to be the error term in the design (within-groups variance or interaction variance) against which to test the significance of main effects or interaction effects. We turn now to this question.

WITHIN-GROUPS VARIANCE AS ERROR

Under the circumstance where the physical conditions of the experiment are selected by nonsampling methods, generalization must be confined to the population of subjects sampled with respect to the *particular* physical conditions present in the experiment. Within-groups variance is legitimate as error, since here is where random sampling took place. Therefore, generalization takes place with regard to subjects only.

But—what about generalizations concerning the conditions of the experiment? Here the crucial issue between systematic factorial design and representative design lies in the *manner of covariation of the physical condition variables among one another*. Unless their *arrangement* (not merely their number) in the experiment is considered in light of the conditions toward which the generalization is aimed, our conclusions are restricted from that generalization.

Fisher was very conscious of the necessity for achieving generalization concerning the range of conditions. For example: "The exact standardization of experimental conditions, which is often thoughtlessly advocated as a panacea, always carries with it the real disadvantage that a highly standardized experiment supplies direct information only in respect of the narrow range of conditions achieved by standardization. Standardization, therefore, weakens rather than strengthens our ground for inferring a like result, when, as is invariably the case in practice, these conditions are somewhat varied" (11, p. 97). Fisher also considered the problem of the arrangement of conditions to be extremely important in multivariate designs and wrote two papers (12, Ch. 17, 28) contrast-

ing the systematic arrangement of field conditions to random arrangement—to the disadvantage of systematic arrangement. Brunswik draws the issue more sharply, pointing out that "generalizability of results concerning . . . the variables involved must remain limited unless at least the range, but better also the distribution . . . of each variable, has been made representative of a carefully defined universe of conditions" (3, p. 53). Brunswik goes beyond Fisher in asserting that systematic factorial designs frequently "tie," or link together, physical condition variables according to the convenience (sometimes arithmetic) of the experimenter's circumstance (3, p. 6). Likewise, variables may be "untied"; that is, the links or correlations among variables in the situation to which the results are to be applied are often disrupted, usually through the "hold all other variables constant," or "isolate a variable" procedure. Note the manner in which certain stimulus variables are arbitrarily "tied" and "untied" in the following example.

Personality and sex. Holtzman (17) was interested in ascertaining the effect of the examiner as a variable in the Draw-A-Person Test. His procedure in establishing the independent variables of sex and of personality was as follows:

Four experienced examiners, two male and two female, were chosen from a group of advanced graduate students in clinical psychology. The two pairs of examiners were selected so as to maximize differences in examiner appearance and personality within both sexes. Examiner M1 was nearly a foot taller and sixty pounds heavier than the other male examiner, M2. The two female examiners, F1 and F2, were approximately the same size but differed considerably in feminine qualities" (17, p. 145).

In line with our general emphasis

on the need for scrutinizing the manner in which the independent variable is constituted in these experiments, note that height and weight were "separated" or varied "within sex" for the male examiners (M1 was a foot taller and sixty pounds heavier than M2). On the other hand, the height of the female examiners was the same (their weights are not mentioned), but their "feminine qualities" "differed considerably." Apparently several different stimulus variables are "tied" (height with female sex) and "untied" (height and weight from male sex and "feminine qualities" within female sex). This situation obscures the independent variable. This obscurity is bound to result when randomization is not effected in the sampling procedures, whether we are dealing with sampling of subjects or "objects" (in this case, examiners).³

Thus, when conditions are not samples, generalization is limited to the subject-population and therefore within-group variance is the only legitimate error term.

INTERACTION TERMS AS ERROR

Now consider the circumstance

³ Note the hypothesis to be tested under these circumstances "(1) The sex of the examiner has a measurable effect . . . (2) The personal characteristics of the examiner aside from sex have a measurable effect . . ." (17, p. 145). The "male variable" is established by drawing two males from the population, one each from some hypothetical personality-type population. The same for the "female variable." From this sample of examiners Holtzman attempts to refute the findings of Sinnett and Eglash (22) (who also used two examiners) concerning the relationship between examiner personality and response to the Draw-A-Person Test. Obviously these particular examiners differed in many variables other than sex characteristics. Such variables can be eliminated as "causes" of results only by randomization—exactly as they are eliminated in sampling subject populations.

where the interaction term might be considered as error. This is the point at which textbooks begin to meet the problem of random sampling of physical conditions. In his discussion of the interaction term as error in his *Design of Experiments* (11, pp. 201–205), Fisher deals with the problem of random sampling of conditions almost to the exclusion of other topics. Both Edwards (10) and Lindquist (18) consider this problem in connection with psychological and educational experiments in which methods, schools, etc. are conditions of the experiment.

Edwards (10, p. 252) in discussing interactions as error terms states, "... it would be illogical to argue that the . . . particular methods . . . selected for investigation have been randomly selected from a population of methods." Lindquist (18, p. 169) discusses this problem through a hypothetical experiment where style and size of type are the independent variables. "... the particular styles (or sizes) involved may not strictly be considered as a random sample from a 'population' of styles (or sizes). The interaction variance in a factorial design is therefore usually not strictly a measure of normally distributed random fluctuations, which theoretically must be true of the error term in any F-test or t-test."

When discussing the necessity for randomness in the use of interaction terms as error terms, Edwards makes very explicit the fact that it is "illogical to argue" that particular methods have been "randomly selected" and that particular instructors cannot be considered random samples. Yet he chooses as an example an experiment by Child (7) which violates this requirement (10, p. 261). Edwards describes the conditions as follows: "The variables introduced were as follows: the sex of the children used

as subjects in the experiment; the sex of the experimenter present during the test situation; the nature of the barrier introduced between the subject and the distant goal object; and the type of instructions given to the child." Child recognizes that his conclusion that "choice of a distant goal was more frequent in the presence of a woman experimenter than in the presence of a man experimenter" is "severely limited by the fact that the sample of experimenters was limited to one of each sex" (7, p. 30). However, Child does use the pooled interaction term as error. "The 11 degrees of freedom pertaining to the 11 possible interactions were therefore pooled for an estimate of error" (7, p. 19). If the writer has correctly identified Child's interaction terms, four ($E \times \text{barrier} \times \text{instructions}$, $E \times \text{barrier}$, $E \times \text{instructions}$, $\text{barrier} \times \text{instructions}$) of the 11 interactions do not include a variable selected at random. As Edwards put it, "furthermore, and this is most important, it is necessary that the categories of one or more of the variables in the experimental design be a random selection from the population being sampled" (10, p. 252). This is followed by a footnote: "This condition will not be met by argument after the experiment has been carried through to completion."

In summary, then, random sampling of the conditions, physical or otherwise, of the experiment is prerequisite to the use of interaction terms as error variance. Ignoring this prerequisite violates the principle of generalization through randomization in exactly the same fashion as in the case of a nonsampling one-factor design.

AGRICULTURAL DESIGNS IN PSYCHOLOGY

The above remarks concerning

interaction terms lead directly to another issue—the use of agricultural designs in psychology. For example, in Lindquist's discussion of interaction terms as error, after very clearly pointing out the logic involved in compromising with the statistical assumptions involved, he concludes the discussion by emphasizing that "... the procedure just recommended is arbitrary in character, *although wide experience in agricultural research indicates that it is usually satisfactory*" (italics ours) (18, p. 170).⁴ Snedecor, however, while discussing interaction terms as error terms in connection with an agricultural experiment, says, "The requirement of randomness is apparently never met in this kind of work, so that statements about probability must be considered inexact" (23, p. 303).

The above remarks concerning agricultural research have been emphasized because they are the key to a fundamental issue. The issue is whether the application of agricultural experimental procedures to psychological experiment needs further scrutiny. We believe that such scrutiny is in order on grounds that there are important discrepancies between the aims and conditions of most agricultural research and most psychological research.⁵ The following points bear on this question.

The type of agricultural research for which Fisher developed factorial design was primarily *engineering-type* research. He was concerned "... that, in any case, there will be no reason for rejecting the experi-

⁴ Lindquist also notes (18, p. 169) that there will be some circumstances where this procedure will not be reasonable.

⁵ See McNemar (19) for a discussion of an important difference in the conditions of agricultural and psychological research in connection with latin-square design.

mental results on the ground that the test was made in conditions differing in one or other of these respects from those in which it is proposed to apply the results" (11, p. 98). In engineering-type experiments the selection and arrangement of the conditions are usually dictated by the particular circumstances to which the experimenter wishes to *apply* his results—and to which he confines his results and conclusions, as Fisher made clear. A further characteristic of engineering-type experiments is that *future* conditions are under the control of the experimenter, so that he can maintain the conditions of the experiment, e.g., issuing the same ratios of ingredients in the ration, the same ratios of chemicals in the fertilizer, etc., as used in the experiment. Therefore, Fisher's factorial design methods are eminently appropriate to the applied problems with which he was concerned. Where the problem of generalizing beyond a particular plot—the soil heterogeneity problem—was concerned, however, he was definite and explicit about the advantages of randomization. (See 11, 12.)

Psychologists, however, have other problems at stake than those of application; for example, it is in the nature of the theoretical task to seek gen-

erality. It seems legitimate, therefore, to ask if psychologists really wish to confine their conclusions to the particular conditions of the experiment, in which the manner of arrangement of the variables is frequently due to convenience, or as Fisher put it, due to the "thoughtless advocacy of standardization as a panacea" (11, p. 97). Moreover, if the future conditions to which the psychologist wishes to predict are not under his control, it also seems legitimate to ask if the manner of covariation of the variables has been maintained in the experiment as in the future situation. If the psychologist does not wish to accept the limitations of systematic design, however, a thorough scrutiny of the principles of representative design is in order, for it is precisely the question of generalization with which representative design is concerned.

SUMMARY

This paper illustrates how unwarranted conclusions may be reached through the application of traditional systematic design to a given problem in clinical psychology—the effect of the examiner on the subjects' responses. Both one-factor and multi-factor designs are discussed.

REFERENCES

1. BRUNSWIK, E. Organismic achievement and environmental probability. *Psychol. Rev.*, 1943, **50**, 255-272.
2. BRUNSWIK, E. Distal focussing of perception. Size-constancy in a representative sample of situations. *Psychol. Monogr.*, 1944, **56**, No. 1 (Whole No. 254).
3. BRUNSWIK, E. *Systematic and representative design of psychological experiments*. Berkeley: Univer. of California Press, 1947. (Also in J. Neyman [Ed.], *Berkeley symposium on mathematical statistics and probability*. Berkeley: Univer. of California Press, 1949. Pp. 143-202.)
4. BRUNSWIK, E. Note on Hammond's analogy between "relativity and representativeness." *Phil. Sci.*, 1951, **18**, 212-217.
5. BRUNSWIK, E. The conceptual framework of psychology. Chicago: Univer. of Chicago Press, 1952. (*Int. Encycl. unified Sci.*, v. 1, No. 10.)
6. BRUNSWIK, E., & KAMIYA, J. Ecological cue validity of "proximity" and of other Gestalt factors. *Amer. J. Psychol.*, 1953, **66**, 20-32.

7. CHILD, I. L. Children's preference for goals easy or difficult to obtain. *Psychol. Monogr.*, 1946, **60**, No. 4 (Whole No. 280).
8. CURTIS, H. S., & WOLF, E. B. The influence of the sex of the examiner on the production of sex responses on the Rorschach. *Amer. Psychologist*, 1951, **6**, 345. (Abstract)
9. DUKES, W. F. Ecological representativeness in studying perceptual size-constancy in childhood. *Amer. J. Psychol.*, 1951, **64**, 87-93.
10. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
11. FISHER, R. A. *The design of experiments*. (4th Ed.) New York: Hafner, 1947.
12. FISHER, R. A. *Contributions to mathematical statistics*. New York: Wiley, 1950.
13. GARFIELD, S. L., BLEK, L., & MELKER, F. The influence of method of administration and sex differences on selected aspects of TAT stories. *J. consult. Psychol.*, 1952, **16**, 140-144.
14. GIBBY, R. G. Examiner influence on the Rorschach inquiry. *J. consult. Psychol.*, 1952, **16**, 449-455.
15. HAMMOND, K. R. Subject and object sampling—a note. *Psychol. Bull.*, 1948, **45**, 530-533.
16. HAMMOND, K. R. Relativity and representativeness. *Phil. Sci.*, 1951, **18**, 208-211.
17. HOLTZMAN, W. H. The examiner as a variable in the Draw-A-Person Test. *J. consult. Psychol.*, 1952, **16**, 145-148.
18. LINDQUIST, E. F. *Statistical analysis in educational research*. Boston: Houghton Mifflin, 1940.
19. MCNEMAR, Q. On the use of latin squares in psychology. *Psychol. Bull.*, 1951, **48**, 398-401.
20. REISS, B. F., SCHWARTZ, E. K., & COTTINGHAM, ALICE. An experimental critique of assumptions underlying the Negro version of the TAT. *J. abnorm. soc. Psychol.*, 1950, **45**, 700-709.
21. ROBINSON, D., & ROHDE, S. Two experiments with an anti-Semitism poll. *J. abnorm. soc. Psychol.*, 1946, **41**, 136-144.
22. SINNETT, E. R., & EGLASH, A. The examiner-subject relationships as a variable in the Draw-A-Person Test. Paper read at the Midwest. Psychol. Ass., Detroit, May, 1950.
23. SNEDECOR, G. W. *Statistical methods*. (4th Ed.) Ames, Iowa: State Coll. Press, 1946.

Received June 14, 1953.

KOLMOGOROV-SMIRNOV TESTS FOR PSYCHOLOGICAL RESEARCH

LEO A. GOODMAN¹

University of Chicago

In an excellent paper, Moses (13) presents some of the principal non-parametric methods and an intuitive explanation of their rationale, properties, and applicability, with a view to facilitating their use by workers in psychological research. One of the important topics in the field of non-parametric methods is the Kolmogorov-Smirnov statistic. Recent results and tables on this topic have been prepared which contribute toward establishing the Kolmogorov-Smirnov statistic as a standard non-parametric tool of statistical analysis.

We shall present an intuitive explanation of the rationale and uses of the Kolmogorov-Smirnov statistic. A table will be presented which facilitates the use of the Kolmogorov-Smirnov statistic by research workers. Some illustrative examples will also be given.

1. TWO-SIDED TESTS

1.1. *One-Sample Test*

In this section tests of "goodness of fit" will be considered. That is, we shall be concerned with the agreement between the distribution of a set of sample values and a theoretical distribution. Probably the most widely used nonparametric test of "goodness of fit" is the chi-square test (4). However, some evidence has been presented indicating that the test which we shall now describe, the

Kolmogorov-Smirnov test for goodness of fit, may be a better all-around test, when it is applicable, than the chi-square test (1, 12). A concise description of other tests of fit appears in (2).

Suppose that a population is thought to have some completely specified cumulative frequency distribution function, say $F(x)$. That is, for any specified value of x , the value of $F(x)$ is the proportion of individuals in the population having measurements less than or equal to x . The observed cumulative step-function $S_N(x)$ of a sample of N observations (that is, $S_N(x) = k/N$, where k is the number of observations less than or equal to x) is expected to be fairly close to this completely specified distribution function $F(x)$. If it is not close enough, we have evidence that the hypothetical distribution $F(x)$ is not the correct one. As a measure of how far the observed cumulative step-function $S_N(x)$ is from the hypothetical distribution $F(x)$ we use the maximum absolute difference d between $S_N(x)$ and $F(x)$; that is, $d = \max_x |F(x) - S_N(x)|$. When d is large we have evidence that $F(x)$ is not the correct population cumulative frequency distribution function.

Let us first consider the case where $F(x)$ is a continuous cumulative distribution function. If the correct population cumulative distribution is in fact $F(x)$, then the sampling distribution of d is known and has been tabulated (1). For example, if $F(x)$ is the correct population distribution, we find from p. 428 in (1) that the prob-

¹ This paper was prepared in connection with research supported by the Office of Naval Research at the Statistical Research Center, University of Chicago. The author wishes to thank Mr. Herbert David, University of Chicago, for helpful comments.

ability that $d \geq 5/15$ is $Pr\{d \geq 5/15\} = 1 - Pr\{d < 5/15\} = 1 - .945 = .055$, and $Pr\{d \geq 6/15\} = 1 - .989 = .011$ for a sample of $N = 15$ observations.

Hence, in order to test the null hypothesis that $F(x)$ is the correct population distribution using a sample of $N = 15$ observations, the maximum absolute deviation d between the sample cumulative step-function $S_{15}(x)$ and $F(x)$ is computed. The null hypothesis is rejected when $d \geq 5/15$, if the null hypothesis is tested at the .055 level of significance. If the test is at the .011 level of significance, the null hypothesis is rejected when $d \geq 6/15$.

Let us now consider the case where $F(x)$ may be a discontinuous cumulative distribution function. Several articles dealing with the Kolmogorov-Smirnov tests claim that the methods apply only where the chance variable is continuous. This is true only if exact probability statements are required. (The reader will note that exact probability statements cannot be made even if the chi-square statistic is used to test goodness of fit since little is known about the actual sampling distribution of the chi-square statistic for finite sample size N and given $F(x)$. The chi-square statistic becomes approximately distribution-free when the sample size N approaches infinity, but is not distribution-free for finite N .) The inequalities stated in (5, 7) serve to validate the use of the Kolmogorov-Smirnov statistic in the case where $F(x)$ may be discontinuous. From these inequalities we see that when $F(x)$ may be discontinuous, the error obtained will be in the "safe direction" if tables are used which assume that $F(x)$ is continuous. More precisely, the level of significance of a test based on the sampling distribution of d will be no larger than the level of significance of that test when

$F(x)$ is assumed continuous. For example, if $F(x)$, which may be discontinuous, is the correct population distribution function, we find from the tables on p. 428 of (1) that the probability that $d \geq 5/15$ is at most .055, and $Pr\{d \geq 6/15\} \leq .011$ for a sample of $N = 15$ observations. Hence, in order to test the null hypothesis that $F(x)$, which may be discontinuous, is the correct population distribution, the value of d is computed. If the null hypothesis is rejected when $d \geq 5/15$, then the level of significance is, at most, .055. The test will be at no more than the .011 level of significance if the null hypothesis is rejected when $d \geq 6/15$. Hence, if the same test is used as when $F(x)$ was assumed continuous, the level of significance will be no more than .055 (or .011).

The Kolmogorov-Smirnov statistic can also be used to estimate probabilities and obtain confidence bands for the true cumulative distribution function $F(x)$ (1, 7, 10, 15, 16). These confidence bands will be free from any restriction concerning the nature of the function $F(x)$. We need not assume that $F(x)$ is continuous in order to obtain confidence bands for $F(x)$. Let us illustrate this use with a particular set of data. A sample of 15 observations was obtained. If this sample is arranged in order of increasing size, we obtain

1, 2, 2, 2, 2, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5.

The reader may be interested to know how these 15 observations were obtained. A student was asked to list three men whom he liked and three men whom he did not like from among all the men he had known since birth. The number 0 was assigned to the person liked the best, number 1 to the person liked second best, number 2 to the person liked third best, and so on, and 5 was as-

signed to the person disliked the most. He was then asked to place the number 0 on a sheet of paper if he thought that his number 0 person (best-liked person) would be the richest (among the total of six men listed). He was to place the number 1 on the paper if he thought that his number 1 person (second best liked) would be richest, and so on, and the number 5 if he thought that the person he disliked most would be the richest among the six. Hence, a single number from 0 to 5 was obtained from this student. Fifteen such observations were obtained by using fifteen different students as the subjects. The sample cumulative step-function $S_{15}(x)$ for the 15 observations is given in Fig. 1. From the tables on p. 428 in (1) we see that the chance is at least .945 that the maximum deviation between $S_{15}(x)$ and the true cumulative distribution $F(x)$

will be less than $5/15$. Hence, if a band of width $5/15$ is drawn above and below $S_{15}(x)$, we can state with at least "94.5 per cent confidence" that the true cumulative distribution $F(x)$ lies within that band. The 94.5 per cent confidence band for $F(x)$ is illustrated in Fig. 1 by the dotted lines above and below $S_{15}(x)$. Therefore, any number of statements of the following kind may be made simultaneously with at least 94.5 per cent confidence: $F(0) < 1/3$, $F(3) < 2/3$, and $F(4)$ is a number between 26.667 per cent and 93.333 per cent.

Let us consider the null hypothesis that a student is equally likely to choose any one of the six numbers. Then the population cumulative distribution function $F(x)$ would be as given in Table 1. The values of $S_{15}(x)$ and $|F(x) - S_{15}(x)|$ are also given in the table. Since the maximum absolute difference between $F(x)$ and $S_{15}(x)$ is $10/30 = 5/15$, the null hypothesis is rejected at the .055 level of significance.

The reader will note that the significance test which was performed was for a completely specified population cumulative distribution function $F(x)$. In cases where parameters must be estimated from the sample (for example, when the null hypothesis is that the population distribution is normal with unspecified mean and standard deviation, and the mean and standard deviation must first be estimated from the sample), there are no theoretical results at present which give exact critical levels for the Kolmogorov-Smirnov statistic. The distribution of d is not known when certain parameters of the population have been estimated from the sample. It may be expected, however, that the effect of adjusting the population mean and standard deviation to those of the sample will be to reduce the

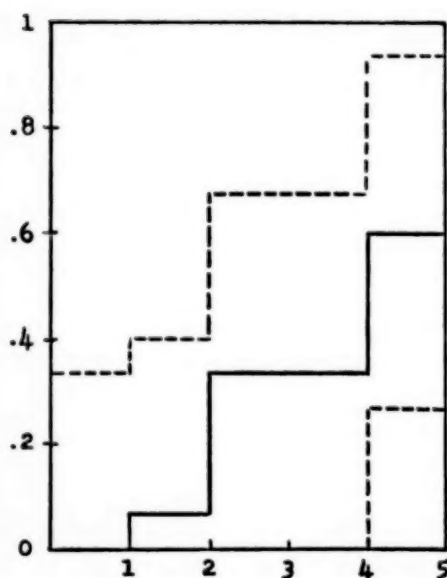


FIG. 1. TWO-SIDED 94.5 PER CENT CONFIDENCE BAND FOR $F(x)$ OBTAINED FROM A SAMPLE OF 15 OBSERVATIONS

TABLE 1

ABSOLUTE DIFFERENCE BETWEEN A SPECIFIED $F(x)$ AND $S_{15}(x)$ FOR A SAMPLE OF 15 OBSERVATIONS

x	0	1	2	3	4	5
$F(x)$	1/6	2/6	3/6	4/6	5/6	1
$S_{15}(x)$	0	1/15	5/15	5/15	9/15	1
$ F(x) - S_{15}(x) $	5/30	8/30	5/30	10/30	7/30	0

critical level of d . If the critical value of d (from tables which assume a completely specified population distribution) is exceeded in these circumstances, we may safely conclude that the discrepancy is significant (see p. 73 of [12]). In cases where parameters must be estimated from the sample, the chi-square test is easily modified by reducing the number of degrees of freedom. The Kolmogorov-Smirnov test has no such known modifications.

1.2. Two-Sample Test

In this section the problem of testing whether two random samples have been drawn from the same population is considered. That is, we shall be concerned with the agreement between the distributions of two sets of sample values.

Let us denote the observed cumulative step-function of the first sample of N observations by $S_N(x)$, and let $S'_M(x)$ be the observed cumulative step-function of the second sample of M observations. The two cumulative step-functions $S_N(x)$ and $S'_M(x)$ are expected to be fairly close to each other if both samples are drawn from the same population. If they are not close enough, we would have evidence that the samples come from different populations. (That is, the population cumulative distribution function for the values from the first sample is different from the population cumulative

distribution for the values from the second sample.) As a measure of how far apart are the two cumulative step-functions we use the maximum absolute difference d' between them; that is, $d' = \max_x |S_N(x) - S'_M(x)|$. When d' is large we have evidence that the samples came from different populations.

Let us first consider the case where the values from both samples are assumed to have continuous population cumulative distributions $F(x)$ and $G(x)$, respectively. We wish to test the null hypothesis that $F(x) \equiv G(x)$ and the null hypothesis will be rejected if the observed value of d' is significantly large. The limiting distribution of d' has been tabled in (15), and a method of obtaining the exact distribution of d' for small samples has been given (11) when in fact $F(x) \equiv G(x)$. A short table for equal size samples is also available (11). The explicit expression for the distribution function of d' has been given recently in (6) for equal size samples. From the tables on p. 126 of (11) we find that, say, in the case where $M = N = 15$, the probability that $d' \geq 7/15$, when in fact $F(x) \equiv G(x)$, is $Pr\{d' \geq 7/15\} = 1 - Pr\{d' \leq 6/15\} = 1 - .925 = .075$, and $Pr\{d' \geq 8/15\} = 1 - .974 = .026$. Hence, in order to test the null hypothesis that $F(x) \equiv G(x)$ at the .075 level of significance, the value of d' is computed and the null hypothesis is rejected when $d' \geq 7/15$. If the test is at the

.026 level of significance, the null hypothesis is rejected when $d' \geq 8/15$.

Let us now consider the case where $F(x)$ and $G(x)$ may be discontinuous cumulative distribution functions. The inequalities stated in (7) serve to validate the use of the Kolmogorov-Smirnov statistic in the case where $F(x)$ and $G(x)$ may be discontinuous. From these inequalities, we see that when $F(x)$ and $G(x)$ may be discontinuous, the error obtained will be in the "safe direction" if tables are used (11) which assume that $F(x)$ and $G(x)$ are continuous. For example, suppose two samples are drawn each containing 15 observations ($M = N = 15$). In order to test the null hypothesis that $F(x) \equiv G(x)$, the value of d is computed. If the null hypothesis is rejected when $d' \geq 7/15$, the level of significance will be, at most, .075. The test will be at no more than the .026 level of significance if the null hypothesis is rejected when $d' \geq 8/15$. The tests will be free from any restriction concerning the nature of the functions $F(x)$ and $G(x)$. We need not assume that the functions $F(x)$ and $G(x)$ are continuous in order to obtain tests of the hypothesis that $F(x) \equiv G(x)$.

Let us illustrate the problem of testing whether two samples have been drawn from the same population. We shall study the agreement between the sample of 15 observations which was described in the preceding section (numbers from 0 to 5 obtained by using fifteen students as

subjects of an inquiry) and a second sample of 15 observations. If the second sample is arranged in order of increasing size, we obtain

0, 0, 0, 0, 1, 1, 2, 2, 2, 2, 3, 3, 5, 5, 5.

(The reader may be interested to know that these 15 observations were obtained by using fifteen businessmen as the subjects of the inquiry. That is, each businessman was interrogated in the same manner as the students. The businessman was then asked to choose the person who would be the richest among the six men listed in the order of his preference.) The values of the cumulative step-functions $S_{15}(x)$ and $S'_{15}(x)$ for the first and second samples respectively are given in Table 2. The values of $|S_{15}(x) - S'_{15}(x)|$ are also given in the table. Since the maximum absolute difference between $S_{15}(x)$ and $S'_{15}(x)$ is $7/15$, the null hypothesis is rejected at the .075 level of significance.

2. ONE-SIDED TESTS

2.1. One-Sample Test

In Section 1.1 the statement was made that the Kolmogorov-Smirnov test for goodness of fit may be a better all-around test, when it is applicable, than the chi-square test. By an all-around test of goodness of fit, we mean a test of the null hypothesis that the observed sample was drawn from a completely specified population without specifying the nature of

TABLE 2
ABSOLUTE DIFFERENCE BETWEEN THE CUMULATIVE STEP-FUNCTIONS FOR TWO SAMPLES
EACH CONTAINING 15 OBSERVATIONS

x	0	1	2	3	4	5
$S_{15}(x)$	0	1/15	5/15	5/15	9/15	1
$S'_{15}(x)$	4/15	6/15	10/15	12/15	12/15	1
$ S_{15}(x) - S'_{15}(x) $	4/15	5/15	5/15	7/15	3/15	0

the alternate hypotheses. That is, the null hypothesis that the observed sample was drawn from a completely specified population is tested against the alternate hypothesis that it was not drawn from that population. In some particular problems more specific alternate hypotheses may be desirable. For example, very often we want to decide not whether an experimental group is the same or different from the general population, but whether the experimental group is better than the general population, or more adjusted than the general population, etc. In such cases, the null hypothesis that there is no difference would be tested against the alternate hypothesis that the experimental group is better, or more adjusted, etc.

Let us consider the numerical illustration presented in Section 1.1 where the null hypothesis that a student is equally likely to choose any one of the six numbers 0, 1, 2, 3, 4, 5 was tested against the alternate hypothesis that the student is not equally likely to choose any one of the six numbers. For this particular problem and for the student body which was under investigation it seemed reasonable to expect that either (a) the six numbers would be equally likely or (b) there would be a tendency for the students to assign the higher numbers. Hence, in this case it is desirable to test the null hypothesis that the six numbers were equally likely against the alternate hypothesis that there was a tendency for the students to assign the higher numbers.

Let us try to make more precise the statement that "there would be a tendency for the students to assign higher numbers." If the six numbers were equally likely, then the proportion of the population assigning the number 0 would be $F(0) = 1/6$. Also

the proportion assigning the number 0 or 1 would be $F(1) = 2/6$, and the proportion assigning numbers no larger than 2 would be $F(2) = 3/6$. Similarly, $F(3) = 4/6$, $F(4) = 5/6$, and $F(5) = 1$. If "there is a tendency to assign higher numbers," then the proportion $G(0)$ of the population assigning the number 0 would be no more than $1/6$ (the case where all numbers are equally likely). Also the proportion $G(1)$ assigning the number 0 or 1 would be no more than $2/6$, and the proportion $G(2)$ assigning the numbers no larger than 2 would be no more than $3/6$. Similarly $G(3) \leq 4/6$, $G(4) \leq 5/6$, and $G(5) \leq 1$. Hence, the statement, "there would be a tendency for the student to assign higher numbers," may be replaced by the statement that the true population cumulative distribution function $G(x)$ is no more than $F(x)$; that is, $G(x) \leq F(x)$ for all values of x .

The null hypothesis that the true population cumulative distribution is, in fact, the specified $F(x)$ is to be tested against the alternate hypothesis that the true population cumulative $G(x)$ is no more than $F(x)$ (with the "less than" relation holding for some values of x). This problem may be considered the one-sided analog of the problem discussed in Section 1.1 where the null hypothesis that the true population cumulative distribution is in fact the specified $F(x)$ is tested against the alternate hypothesis that the true population cumulative is not $F(x)$. As a measure of how far the observed cumulative step-function $S_N(x)$ is from the hypothetical distribution $F(x)$ we use the maximum difference c between $S_N(x)$ and $F(x)$. That is, $c = \text{maximum}_x [F(x) - S_N(x)]$, which is the one-sided analog of the maximum absolute difference d . When c is large, we have evidence that $F(x)$ is not the correct population cumulative distribution

and that the true population cumulative $G(x)$ is no more than $F(x)$. If the correct population cumulative is continuous and is in fact $F(x)$, then the sampling distribution of c is known, and the explicit expression for the distribution function of c is given by equation 3.0 on p. 593 in (3). Using equation 3.0, we find for example that when $N=15$ the chance that c will be no more than $5/15$ is

$$P_{15}(5/15) = 1 - \frac{1}{3} \sum_{j=0}^{10} \binom{15}{j} \left(\frac{2}{3} - \frac{j}{15} \right)^{15-j} \left(\frac{1}{3} + \frac{j}{15} \right)^{j-1} = .97.$$

Hence, the probability that $c \geq 5/15$ is $Pr\{c > 5/15\} = 1 - .97 = .03$. In order to perform a "one-sided test" of the null hypothesis that $F(x)$ is the correct population distribution using a sample of $N=15$ observations, the value of c is computed. The null hypothesis is rejected when $c \geq 5/15$ if the test is at the .03 level of significance.

For the particular numerical illustration presented in Section 1.1, we find that $c = 5/15$. Hence, the null hypothesis that a student is equally likely to choose any one of the six numbers is rejected at the .03 level of significance, and the alternate hypothesis that there is a tendency to assign higher numbers is accepted.

In Section 1.1 a method was given for obtaining two-sided confidence bands for the true cumulative distribution function $F(x)$. This method may be modified in order to obtain one-sided confidence bands for $F(x)$. For example, Fig. 1 gives a two-sided 94.5 per cent confidence band for $F(x)$ obtained from a sample of 15 observations. We might make the one-sided confidence statement that the true cumulative distribution $F(x)$ lies below the upper limit of the band

presented in Fig. 1. From the results presented earlier in this section we see that there is at least "97 per cent confidence" that $F(x)$ lies below the upper limit of the band in Fig. 1.

2.2. Two-Sample Test

In this section a one-sided analog of the problem discussed in Section 1.2 will be considered. We wish to test the null hypothesis that the cumulative distribution function $F(x)$ for the values from the first sample is equal to the cumulative distribution function $G(x)$ for the values from the second sample. In other words, the null hypothesis is that $F(x) \equiv G(x)$. The alternate hypothesis to be considered is of a more specific nature than the alternate hypothesis for the all-around test presented in Section 1.2. We shall be concerned with the alternate hypothesis that $F(x) \leq G(x)$. If the alternate hypothesis is, in fact, true (with the "less than" relation holding for some values of x), we say that the population values from which the first sample was drawn are stochastically larger than the population values from which the second sample was drawn. The importance of such alternate hypotheses has been stressed in (8, 9). For example, very often we want to decide not whether the experimental group is the same or different from the control group but whether the experimental group is better than the control group, or more adjusted than the control group, etc. In such cases the null hypothesis that $F(x) \equiv G(x)$ would be tested against the alternate hypothesis that $F(x) \leq G(x)$. For this one-sided analog of the problem described in Section 1.2, we use the maximum difference c' between the observed cumulative step-functions $S_N(x)$ and $S'_M(x)$ of the first sample of N observations and of the second sample of M observations. That is, $c' = \text{maximum}_x$

$[S'_M(x) - S_N(x)]$ which is the one-sided analog of the maximum absolute difference d' . When c' is large, we have evidence that $F(x)$ is not equal to $G(x)$ and that the population values from which the first sample was drawn are stochastically larger than the population values from which the second sample was drawn. If the population cumulative functions are continuous and in fact $F(x) \equiv G(x)$, then the limiting distribution of c' is known (9, 14). If $F(x) \equiv G(x)$, we find that the sampling distribution of $4(c')^2 MN/(M+N)$ will have approximately a chi-square distribution with two degrees of freedom when M and N are large and M/N is not too close to either zero or infinity. Hence, the tables of the chi-square distribution may be utilized to test the null hypothesis when M and N are large. When M and N are small, the exact distribution of c' may be computed by extending the counting method presented in (11) for the two-sided problem. The explicit expression for the distribution function of c' has been given recently (6) for equal-size samples. Table 3 may be used to test the null hypothesis at either the 20 per cent, 10 per cent, 5 per cent, 1 per cent, or 0.1 per cent level of significance if $M=N$. The table gives the critical value of $c'N$ at the various levels of significance. For example, when $N=15$ we see from

Table 3 that 7 is the critical value of $c'N$ at the 5 per cent level of significance. Hence, the null hypothesis is rejected at the 5 per cent level of significance if $c' \geq 7/15$. Using the explicit expression (see [6]) for the distribution function of c' , we find that

$$\begin{aligned} \Pr \{c' \geq 7/15\} &= 1 - \Pr \{c' < 7/15\} \\ &= \binom{30}{8} / \binom{30}{15} = .038. \end{aligned}$$

Let us reconsider the numerical il-

TABLE 3
CRITICAL VALUES OF $c'N$ FOR THE $\alpha \cdot 100$
PER CENT LEVEL OF SIGNIFICANCE

$N \backslash \alpha$.001	.01	.05	.1	.2
1	∞	∞	∞	∞	∞
2	∞	∞	∞	∞	2
3	∞	∞	3	3	3
4	∞	∞	4	4	3
5	∞	5	4	4	3
6	∞	6	5	4	4
7	7	6	5	5	4
8	8	6	5	5	4
9	8	7	6	5	4
10	9	7	6	5	5
11	9	8	6	6	5
12	9	8	6	6	5
13	10	8	7	6	5
14	10	8	7	6	5
15	10	9	7	6	5
16	11	9	7	7	6
17	11	9	8	7	6
18	11	10	8	7	6
19	12	10	8	7	6
20	12	10	8	7	6
21	12	10	8	7	6
22	13	11	9	8	6
23	13	11	9	8	7
24	13	11	9	8	7
25	13	11	9	8	7
26	14	11	9	8	7
27	14	12	9	8	7
28	14	12	10	9	7
29	14	12	10	9	7
30	15	12	10	9	7
35	16	13	11	9	8
40	17	14	11	10	9
45	18	15	12	11	9
50	19	16	13	11	9

lustration presented in Section 1.2. Since $c' = 7/15$, we reject the null hypothesis at the .038 level of significance that the population distribution of the numbers assigned by the students was the same as the population distribution of the numbers assigned by the businessmen and accept the alternate hypothesis that the numbers assigned by the students were stochastically larger than those assigned by the businessmen. In other words, we accept the alternate

hypothesis that there was a greater tendency for the students to assign higher numbers than the businessmen.

It is interesting to note that if approximate critical values are computed using the chi-square (with two degrees of freedom) approximation the approximate critical values are always more than the exact critical values (in Table 3) minus 1. Hence, the error in using the chi-square approximation is always in the "safe direction" for the levels of signifi-

cance in Table 3 even when the sample size is small. In other words, if the null hypothesis is rejected using the chi-square approximation, it would also be rejected if exact computations had been made. It is also interesting to note that only in the following cases will the chi-square approximation lead to acceptance when an exact computation based on Table 3 would lead to rejection: $\alpha = .001$ and $N = 12, 15, 18, 21, 25, 29$; $\alpha = .01$ and $N = 8, 14$.

REFERENCES

1. BIRNBAUM, Z. W. Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample sizes. *J. Amer. statist. Ass.*, 1952, **47**, 425-441.
2. BIRNBAUM, Z. W. Distribution-free tests of fit for continuous distribution functions. *Ann. math. Statist.*, 1953, **24**, 1-8.
3. BIRNBAUM, Z. W., & TINGEY, F. H. One-sided confidence contours for probability distributions functions. *Ann. math. Statist.*, 1951, **22**, 592-596.
4. COCHRAN, W. G. The χ^2 test of goodness of fit. *Ann. math. Statist.*, 1952, **23**, 315-345.
5. DAVID, H. T. Discrete populations and the Kolmogorov-Smirnov tests. Unpublished report SRC-21103D27, Statist. Res. Cent., Univer. of Chicago.
6. GNEDENKO, B. V., & KOROLYUK, V. S. On the maximum discrepancy between two empirical distributions. *Doklady Akad. Nauk S.S.S.R. (N.S.)*, 1951, **80**, 525-528. (A review of this article appears in *Mathematical Reviews*, 1952, **13**, 570.)
7. KOLMOGOROV, A. Confidence limits for an unknown distribution function. *Ann. math. Statist.*, 1941, **12**, 461-463.
8. MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50-60.
9. MARSHALL, A. W. A large-sample test of the hypothesis that one of two random variables is stochastically larger than the other. *J. Amer. statist. Ass.*, 1951, **46**, 366-374.
10. MASSEY, F. J., JR. A note on the estimation of a distribution function by confidence limits. *Ann. math. Statist.*, 1950, **21**, 116-119.
11. MASSEY, F. J., JR. The distribution of the maximum deviation between two sample cumulative step-functions. *Ann. math. Statist.*, 1951, **22**, 125-128.
12. MASSEY, F. J., JR. The Kolmogorov-Smirnov test for goodness of fit. *J. Amer. statist. Ass.*, 1951, **46**, 68-78.
13. MOSES, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122-143.
14. SMIRNOV, N. Sur les écarts de la courbe de distribution empirique. *Recueil Mathématique (Mathematiceskii Sbornik, M.S.)*, 1939, **48**, 3-26.
15. SMIRNOV, N. Table for estimating the goodness of fit of empirical distributions. *Ann. math. Statist.*, 1948, **19**, 279-281.
16. WALD, A., & WOLFOWITZ, J. Confidence limits for continuous distribution functions. *Ann. math. Statist.*, 1939, **10**, 105-118.

Received April 1, 1953.

REMARK ON "A QUALIFICATION IN THE USE OF ANALYSIS OF VARIANCE"

VICTOR H. DENENBERG

Human Resources Research Office, Fort Knox, Kentucky

In their original article (7) Webb and Lemmon stated that the results of the over-all F test may not agree with results obtained by subsequent use of the t test applied to individual means when a functional relationship exists between the independent and dependent variables. They gave hypothetical examples which, they thought, illustrated their point. Patterson (5) and Diamond (1) have both taken exception to the original article, but Webb and Lemmon in their reply (8) stated that "... most of their [Patterson and Diamond's] criticisms have not been aimed directly at the core of our problem. Most of their discussion seems to deal with problems of the conventional analysis of variance situation, in which the means of the groups show no trend, but are randomly related to each other. . . . We were concerned, however, with a fairly common experimental design in which the means of the groups show a definite ordering or trend, with respect to some other experimental variable."

In all four of the articles written by the above authors, analysis of variance has been discussed as though its only use were to test for significance of differences between means. This is probably what analysis of variance has been most often used for in psychological research, but it has other uses which are just as important as the test of differences between means. One of the other uses of analysis of variance is that of testing to see whether there is a *significant linear*

or curvilinear regression between the independent and dependent variables. This is the test which should be used with Webb and Lemmon's data since they hypothesize that a functional relationship exists. This test may be performed by an extension of the conventional analysis of variance procedure. Fisher's (2) method of orthogonal polynomials is convenient to use for this analysis. Snedecor (6, Ch. 14, 15) discusses in detail the procedures necessary to make this analysis, and Johnson and Tsao (3) give a very complete discussion of the procedures involved in fitting orthogonal polynomials to a $4 \times 7 \times 2 \times 2 \times 2$ factorial design on determination of differential limens.

The purpose of this note is to comment briefly on this technique and show how its use resolves the seeming paradox posed by Webb and Lemmon. The principle underlying this procedure is that, when a test of functional relationship is desired, the between SS with $k-1$ df may be analyzed into $k-1$ orthogonal comparisons, each associated with a particular aspect of the regression line. Each of these components may be tested for significance independently. The first component isolated is that due to linearity, the second is the quadratic component, the third the cubic, etc. The df associated with each of these is 1, and the individual MS 's are tested against the within MS for significance.

Cases I, II, and V from Webb and Lemmon's original article will be discussed with reference to this

method. Cases III and IV have to be eliminated since it is necessary, when using orthogonal polynomials, either to assume that observations are taken at equal intervals along the X -axis, or else the actual values on the X -axis must be known. For Case I there is no problem since only two sets of observations are taken. In Case II Webb and Lemmon assume that their C group is midway between groups A and B, so equal intervals are present here. For Case V inspection of the graph seems to indicate that the four groups are spaced at equal intervals, and that is the assumption made here. For Cases III and IV it is obvious that the groups are not spaced at equal intervals, so these cannot be discussed.

To illustrate the procedure and its interpretations, the various groups were assigned arbitrary summation values (the results would have been the same no matter what values were assigned), and analysis of variance tables were constructed which yield F ratios identical with those given by Webb and Lemmon. For Case I, ΣA was assumed to be 11 and ΣB was

assumed to be 33. With 11 Ss in each group it was simple to determine the SS and MS for the between-groups variance. Since $F = 4.35$ by Webb and Lemmon's definition, it was possible to work back and determine what the within MS and SS had to be to yield this F ratio. The same procedure was followed for Cases II and V. For Case II, $\Sigma A = 11$, $\Sigma B = 33$, and $\Sigma C = 22$. For Case V, $\Sigma A = 11$, $\Sigma B = 11$, $\Sigma C = 33$, and $\Sigma D = 33$. Then the between SS for Case II was analyzed into its linear and quadratic components, and the between SS for Case V was analyzed into its linear, quadratic, and cubic components. The MS 's were then computed and were tested against the within MS . All these data are illustrated in Table 1. The values which are not in parentheses are the between and within sources of variance and are identical with Webb and Lemmon's data. All the data which are enclosed in parentheses are the results of the orthogonal polynomial analysis.

When Case II is analyzed in this manner, it is seen, though the overall F test is insignificant, that when

TABLE 1
ANALYSIS OF VARIANCE OF CASES I, II, AND V FROM WEBB AND LEMMON'S
HYPOTHETICAL DATA (7)

Data not in parentheses are from Webb and Lemmon. Data in parentheses are the results of the orthogonal polynomial analysis.

Case	Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
I	Between	1	22	22	4.35
	Within	20	101.20	5.06	
II	Between	2	22	11	2.17
	(Linear)	(1)	(22)	(22)	(4.35)
	(Quadratic)	(1)	(0)	(0)	
	Within	30	151.80	5.06	
V	Between	3	44	14.67	2.90
	(Linear)	(1)	(0)	(0)	(8.70)
	(Quadratic)	(1)	(44)	(44)	
	(Cubic)	(1)	(0)	(0)	
	Within	40	202.40	5.06	

the linear and quadratic components are separated all the between variance is explained by the linear regression (which would have to be true since Webb and Lemmon assumed that the mean of their C group fell directly on the grand mean). Since there are 1 and 30 *df* to test the linear *MS*, this linear regression is more significant than that for Case I where the identical *F* ratio is found but is tested against 1 and 20 *df*. (Case I may also be considered to be a test of linearity, since the linear *SS* here is identical with the between *SS*.) This also follows logically since we would place more confidence in a linear curve determined by 3 points than we would in a curve determined by only 2 points.

For Case V it may be seen that a highly significant quadratic relationship exists between the two variables. None of the regression is explained by the linear or cubic elements; the quadratic component accounts for all the between variance. Though in this case the over-all *F* is found to be barely significant at the .05 level, a more accurate interpretation of the experiment is obtained from this further analysis.

This type of statistical analysis

would appear to be profitable in psychological research. As Webb and Lemmon indicate, the situation where a functional relationship exists between the independent and dependent variables is a fairly common one. Kogan (4) in his recent review of variance designs in psychological research points out that the procedure of fitting orthogonal polynomials will frequently furnish information and answers to questions which cannot be obtained by an over-all *F* test to treatment means. However, the fact that Kogan only cites one psychological study which has used this method—that by Johnson and Tsao (3)—would indicate that this technique has not been widely used by psychologists.

In summary, the author would like to point out that a logical analysis of the experimental situation must precede the selection of a statistical technique. In the case where *E* suspects that a functional relationship exists, the over-all *F* test of treatment means is not the best test of this relationship. The use of the method of orthogonal polynomials, however, will permit *E* to make an exact test of his hypothesis.

REFERENCES

1. DIAMOND, S. Comment on "A qualification in the use of analysis of variance." *Psychol. Bull.*, 1952, **49**, 151-154.
2. FISHER, R. A. *Statistical methods for research workers*. (11th Ed.) New York: Hafner, 1950.
3. JOHNSON, P. O., & TSAO, F. Factorial design in the determination of differential limen values. *Psychometrika*, 1944, **9**, 107-144.
4. KOGAN, L. S. Variance designs in psychological research. *Psychol. Bull.*, 1953, **50**, 1-40.
5. PATTERSON, C. H. Note on "A qualification in the use of analysis of variance." *Psychol. Bull.*, 1952, **49**, 148-150.
6. SNEDECOR, G. W. *Statistical methods*. Ames, Iowa: Collegiate Press, 1946.
7. WEBB, W. B., & LEMMON, V. W. A qualification in the use of analysis of variance. *Psychol. Bull.*, 1950, **47**, 130-136.
8. WEBB, W. B., & LEMMON, V. W. A sequel to the notes of Patterson and Diamond. *Psychol. Bull.*, 1952, **49**, 155.

Received March 5, 1953.

TEST OF SIGNIFICANCE FOR A SERIES OF STATISTICAL TESTS

JAMES M. SAKODA, BURTON H. COHEN, AND GEOFFREY BEALL

University of Connecticut

The problem of evaluating a series of statistical tests (e.g., t 's, F 's, χ^2 's) has recently received the attention of psychologists. The general approach to the problem is to set the significance level (p) at .05 or .01 and to find the chance probability of obtaining at least n significant results. This is done by expanding the binomial, $(p+q)^N$, where $q=1-p$ and N is the number of significance tests made. This procedure is applicable when prediction of the outcome of a statistical test is independent of any other tests in the series (8). Wilkinson (14) has published tables for p at the .05 and .01 levels showing the probability of obtaining n or more significant statistics out of N calculated statistics. Wilkinson's tables, however, only run to $N=25$. Brožek and Tiede (1) in a subsequent article suggested that when Np is equal to or larger than five it is possible to employ the normal curve approximation to the binomial distribution. To do this, the critical ratio is calculated by means of the formula

$$CR = \frac{|n - M| - .5}{\sigma}$$

where $M=Np$, $\sigma=\sqrt{Npq}$, and $-.5$ is the correction for discontinuity. The probability value is obtained from the table of area under one tail of the normal curve. To use this normal curve approximation, however, N must be 100 or larger when p is taken as .05, or 500 or larger when p is taken as .01, since it will be remembered that Np must be equal to or larger than five. Between

Wilkinson's tables and the normal curve approximation there is a gap which we feel should be filled in.

The graphs which we provide here run to $N=100$ for $p=.05$ (Fig. 1) and $N=500$ for $p=.01$ (Fig. 2) and can be used when the normal curve approximation is not applicable. To use the graphs, the .05 or the .01 level of confidence is selected and the number of calculated statistics (N) and the number of significant statistics (n) at the chosen level of confidence are counted. The chance probability of obtaining at least n out of N statistics can be read off the graph for values between .001 and .50. N has been plotted on a logarithmic scale, and this fact should be taken into account in interpolating for values of N . For example, for $n=7$, $N=60$, and $p=.05$ chance probability can be read from Fig. 1 as lying between .05 and .01. One would conclude that it is not probable that obtaining seven significant results out of 60 was due to chance alone. On the other hand, there is still the possibility that several of the seven significant statistics might have occurred by chance alone.

The current practice of tabling critical values of t 's, F 's, χ^2 's at the .05 and .01 levels of confidence makes the counting of significant statistics at these levels and the use of the binomial distribution the most suitable approach to the problem of testing the significance of a series of statistical tests. However, there are two types of situations in which our graphs will be inadequate. The first is one in which the level of signifi-

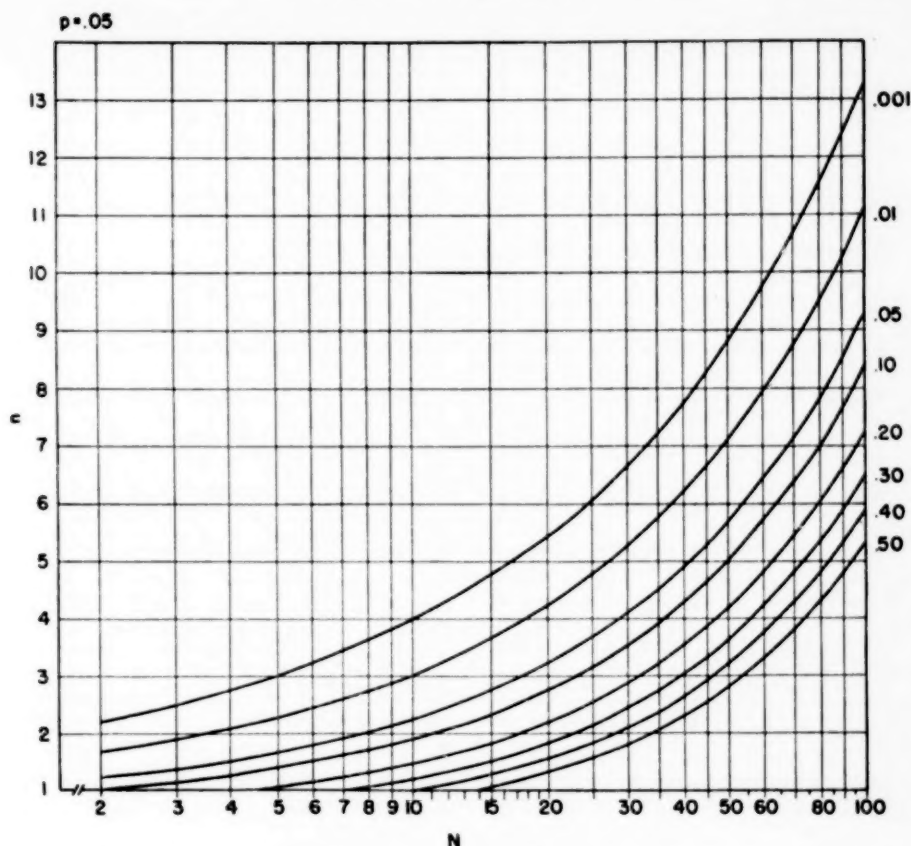


FIG. 1. CHANCE PROBABILITY OF OBTAINING AT LEAST n STATISTICS SIGNIFICANT AT THE .05 LEVEL FROM N CALCULATED STATISTICS

cance one desires to adopt is not .05 or .01, but, for example, .10 or .001. There are at least three ways of handling this situation. The first is to consult tables of binomial distributions. The U. S. National Bureau of Standards (16) has published tables for p ranging from .01 to .50 and N 's up to 50, and Romig (12) has published tables for N 's from 50 to 100. A second method is to calculate the desired binomial distribution directly, using a calculating machine and a convenient working formula (9, pp. 22-23). A third method is to use the Poisson distribution as an approxi-

mation to the binomial distribution (4). When Np is taken to be less than five (the range within which the normal curve approximation is not applicable) and p is taken to be not larger than .10, the approximation of the binomial distribution by the Poisson is fairly good even for values of N as small as two. We have found that with these restrictions the largest absolute error in calculating cumulative probabilities is .02, and in the critical area of cumulative probabilities of .10 or less the error is not larger than .012. Soper (13), Molina (10), and Hartley and Pearson (7)

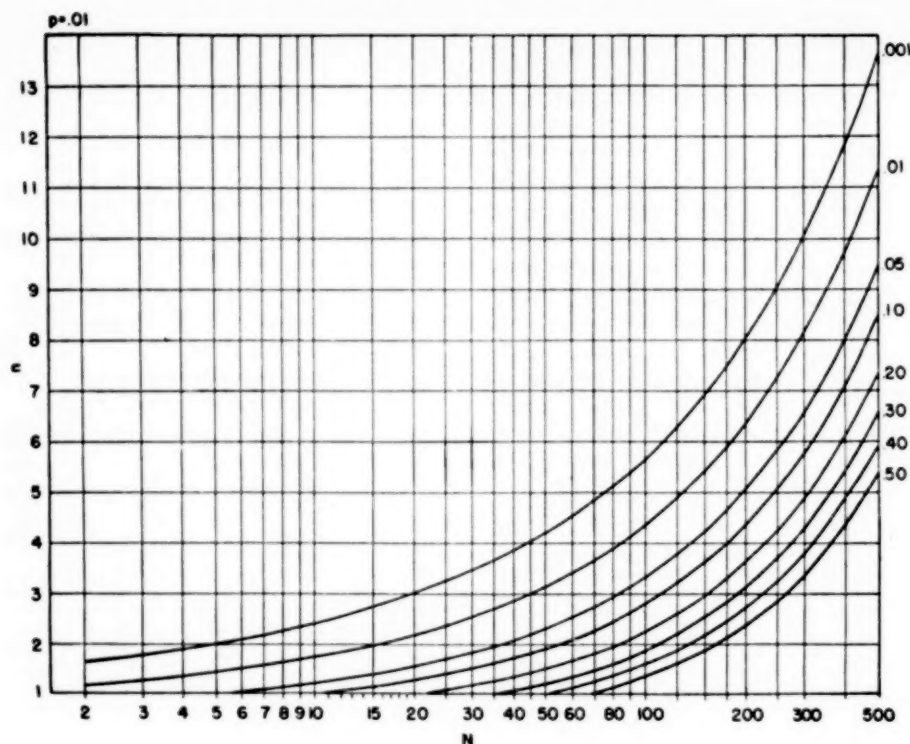


FIG. 2. CHANCE PROBABILITY OF OBTAINING AT LEAST n STATISTICS SIGNIFICANT AT THE .01 LEVEL FROM N CALCULATED STATISTICS

have published tables of Poisson distributions, and Dixon and Massey (3) include an abridged Poisson distribution in the appendix of their book.

A second type of situation in which our graphs will not be adequate is one in which exact probabilities are calculated for a number of significance tests and a sensitive test of over-all significance of the series is desired. Fisher (5) offers a method of combining exact probabilities from a series of tests of significance. His test is based on the formula for the probability of a chi square with two degrees of freedom. Using this formula it is possible to transform p values to chi-square values. Using common logarithms,

$$\chi^2 = 2 \cdot 2.302585(-\log_{10} p).$$

Independent chi squares and their degrees of freedom (two for each chi square) can be summed and these sums referred to a chi-square table for a combined probability value. For an example which is worked out, interested readers are referred to Fisher (5, pp. 99-101) or to the article by Jones and Fiske (8).

To use this method exact probabilities of t 's, χ^2 's, and F 's are needed. Good approximations of exact probabilities can be obtained by linear interpolation in available tables of t and χ^2 by first expressing probabilities in natural or common logarithms. Since Fisher's chi-square technique calls for exact probabilities expressed in logarithms, the values found by interpolation need not be transformed to

their antilogarithm equivalents. The same procedure applies to F , with the additional step of transforming F to square root of F before making the linear interpolation. The usual published tables of F do not allow for interpolations for p values below .05. However, Hald (6) in his *Statistical*

Tables and Formulas includes F distributions for critical p values of .10, .30, and .50. Exact p values for F 's can also be calculated from the incomplete beta function tabled by Pearson (11), and readers are referred to Burke's (2) article explaining this procedure.

REFERENCES

1. BROŽEK, J., & TIEDE, K. Reliable and questionable significance in a series of statistical tests. *Psychol. Bull.*, 1952, **49**, 339-341.
2. BURKE, C. J. Computation of the level of significance in the F test. *Psychol. Bull.*, 1951, **48**, 392-397.
3. DIXON, W. J., & MASSEY, F. J., JR. *Introduction to statistical analysis*. New York: McGraw-Hill, 1951.
4. FELLER, W. *An introduction to probability theory and its applications*. I. New York: Wiley, 1950.
5. FISHER, R. A. *Statistical methods for research workers*. New York: Hafner, 1948.
6. HALD, A. *Statistical tables and formulas*. New York: Wiley, 1952.
7. HARTLEY, H. O., & PEARSON, E. S. Tables of the χ^2 -integral and of the cumulative Poisson distribution. *Biometrika*, 1950, **37**, 313-325.
8. JONES, L. V., & FISKE, D. W. Methods for testing the significance of combined results. *Psychol. Bull.*, 1953, **50**, 375-382.
9. KENNY, J. F. *Mathematics of statistics*, II. New York: Van Nostrand, 1941.
10. MOLINA, E. C. *Poisson's exponential binomial limit*. New York: Van Nostrand, 1949.
11. PEARSON, K. (Ed.) *Tables of the incomplete beta-function*. London: Biometric Lab., University College, 1934.
12. ROMIG, H. G. *50-100 binomial tables*. New York: Wiley, 1952.
13. SOPER, H. E. Tables of Poisson's exponential binomial limit. *Biometrika*, 1914, **10**, 25-35.
14. WILKINSON, B. A. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, **48**, 156-158.
15. *Tables of the binomial probability distribution*. National Bureau of Standards, Applied Mathematics Series 6, Washington, D. C.: U. S. Government Printing Office, 1950.

Received August 27, 1953.

COMMENTS ON SEEMAN'S OPERATIONAL ANALYSIS OF THE FREUDIAN THEORY OF DAYDREAMS

RICHARD A. BEHAN AND FRANCES L. BEHAN

Michigan State College

While the present writers are in wholehearted agreement with Dr. Seeman's remark that formal operational analysis is an indispensable prerequisite to empirical investigation, it is also true that the results of the analysis, once it is completed, need to be criticized. After one has abstracted the formal structure of a theory, it is necessary to subject this formal structure to logical criticism.

The first criticism is concerned with the restatement of the Freudian assertion that daydreams are wish-fulfillments. The actual restatement is as follows: "*The emission of a daydream is functionally related to a specific type of demand (wish), the relation being such that whenever an instance of such and such a daydream is observed, it is required by the theory that an instance of a specified corresponding demand (wish) must be identified by a suitable objective operation*" (2, p. 377). Dr. Seeman then goes on to say: "It seems clear that, so stated, the theory really requires the occurrence of identifiable, lawful patterns of *demand-daydream* covariation." The point to be made here is that this restatement is not a statement in theory at all; it is, rather, metatheory. That is to say, the restatement is a statement *about* theory, not a statement *in* theory (4). If the quoted statement were actually theory, it would be an assertion of a *specific* functional relation. *There are many different functional relations which fit the description given in the quoted statement.*

Second, the theory does not require "the occurrence of identifiable, law-

ful patterns of *demand-daydream* covariation." The theory is the set of statements which assert "identifiable, lawful patterns of demand-daydream covariation." Every empirical theory consists of statements which assert identifiable, lawful patterns of construct phenomena covariation. This is required *of* the theory, not *by* the theory on purely methodological considerations.

Third, on page 377, Dr. Seeman asserts: "What is crucially important here is the understanding of the *contingent notion of frequency*, which lies buried in this analysis of the *meaning* of the concept of wish-fulfillment." Then on page 379, he asserts that Q —the hypothesis that is predicted by the theory—is a predicate. Now, a predicate, with its argument, is a two-valued constant (not a many-valued functor); it is either true or it is false. It is not possible, with the predicate Q analyzed as it is, to predict anything about the frequency of daydreams.

It is well to add in passing that Dr. Seeman is not the only psychologist who has confused statements about theory for statements in theory in this respect. There is, as it were, a great deal of precedent in modern psychological "theorizing" for calling statements of this sort theory.

The next point for consideration concerns the inductive leap from $(P \supset Q) \cdot Q$ to (P) . According to footnote 7 (2, p. 378) the symbolized statement is "an extremely elementary application of symbolic logic." This is not the case. The symbolized statement is an extremely elementary

operation *excluded* in symbolic logic—known as the fallacy of asserting the consequence (1, pp. 7–8).

The reasons that we wish to exclude formal fallacies from our theories may be summed up as follows: A fallacy is a false form of statement. It is a well-known theorem in symbolic logic that a false statement implies any statement (3, p. 104). Now, it is the case that statements asserting both sides of every question about which any theory makes an assertion are included in the class of all statements (the class including every statement). Therefore, the theory which contains even one fallacy will predict both sides of every question about which it contains an assertion. There are three results of this state of affairs which are worth mentioning: (a) The theory is never wrong—it always predicts what is in fact the case, along with what is not the case. (b) The theory makes no unequivocal assertion, and the theorist must always wait until after the fact to find out what his theory would predict. Thus, prediction is always of an *ad hoc* nature. (c) The theory is useless. Any procedure which purports to be the result of the use of the theory could proceed just as well without it.

The next point for consideration concerns the discussion of the sentence $[(P \cdot P'_{1...k}) \supset Q_{1...k}] \cdot [Q]$.

On page 379 Dr. Seeman states "... in those isolated instances ... where the confirmation conditions indicate that $\sim Q$ is the case, the indication would be for a re-examination of P' before P ." With this point of view we disagree. Whenever a statement derived from an empirical theory is disconfirmed the theory itself is denied. The only way out is to show that the theory does not predict the state of affairs that was disconfirmed.

With the help of symbolic logic it is

easy to show that it is P that is false when one discovers a disconfirmation of the theory. There are two cases, namely: (a) P' asserts a fact; (b) P' asserts an assumption (2, p. 379). Consider the first case: If it were true that P' asserts a fact, then P' cannot be false. After all, if P' asserts an actual state of affairs, it can only be true. The theory asserts that $(P \cdot P')$ implies Q . We find $\sim Q$; therefore, by Modus Tollens, we deduce $\sim(P \cdot P')$. But P' is true; therefore, P must be false (1, p. 25).

Consider the second case: P' asserts an assumption; i.e., P' is an instance of P . The logical model for this situation is the reference formula known as Specification (Spec., see 1, p. 354). Applied to our situation Spec. asserts that P is a universal assertion that implies P' . It goes almost without saying that the universe of discourse here is the circumscribed universe in which the theory under consideration applies. Since P implies P' , and we have $\sim P'$, we may, by the reference formula known as Modus Tollens, deduce $\sim P$.

It is thus seen that the occurrence of a nonconfirmation ($\sim Q$) leads in every case to the denial of the theory. *There is no way to save a theory if it actually predicts wrong; one can only change it.*

The present writers hope that the reader will not feel that they disprove of what Dr. Seeman has tried to do. On the contrary, it is felt that Dr. Seeman has accomplished two important things: (a) He has provided the first (to our knowledge) attempt to demonstrate the logical form of the Freudian theory of daydreams. (b) He has stated the results of his analysis in such a way that one can unequivocally determine its logical characteristics.

The present comments will be closed with a few words about the

implication of Dr. Seeman's analysis for the Freudian theory of daydreams. If Dr. Seeman's analysis is correct, then all of the remarks which were made earlier, with reference to theories which contained fallacies, are applicable to the Freudian theory

of daydreams. These remarks were: (a) The "theory" predicts both sides of every question. (b) The "theory" is of an *ad hoc* nature. (c) The "theory" is useless; any procedure which purports to follow from the "theory" could proceed just as well without it.

REFERENCES

1. COOLEY, J. C. *A primer of formal logic*. New York: Macmillan, 1949.
2. SEEMAN, W. The Freudian theory of daydreams: an operational analysis. *Psychol. Bull.*, 1951, 5, 369-382.
3. WHITEHEAD, A. N., & RUSSELL, B. *Principia mathematica*. Vol. 1. Cambridge: Cambridge Univer. Press, 1925.
4. WOODGER, J. H. The technique of theory construction. *Int. Encycl. unified Sci.*, 1939, 2, No. 5.

Received March 16, 1953.

PSYCHOLOGICAL BULLETIN
Vol. 51, No. 2, 1954

REPLY TO THE BEHANS

WILLIAM SEEMAN

Mayo Clinic

In connection with the Behan paper, I shall confine myself to a few brief comments.

1. The Behans' doctrinaire statements about theory suggest a confusion of logical fact with methodological decision, especially with the notion of what Reichenbach calls "entailed decisions" (5, pp. 11-15). The impression conveyed in their paper that their remarks on theory are in accord with Woodger (6) is also erroneous. Actually, they are in disagreement with Woodger; and the "specimen theory" he presents (and which he therefore, presumably, considers acceptable as "theory") would be ruled out by the criteria set out in the Behan paper.

2. It is not true that I have committed the fallacy of asserting the consequent. Given "If 'P' then 'Q' and assert 'Q' " it would be asserting the consequent if *and only if* I were to say "assert 'P' to be true on the basis of 'Q.' " But my paper does not say that; what it says specifically is "inductive leap to 'P,' " and this on

the assumption that it would be understood as a convenient shorthand for something like "accept 'P' provisionally as a proposition in the larger theory until such time (if any) that there is sufficient evidence to controvert it and render it useless in the theory." To call this "asserting the consequent" is to say either that all experimental procedure leads to this fallacy or to assert a stricture against all empirical science. And this, I presume, is what MacCorquodale and Meehl have in mind when they write "All scientific hypothesizing is in the invalid 'third figure' of the implicative syllogism" (4, p. 97). What purpose, after all, can an experiment serve if, from the results, one is forbidden to make inferences?

3. With respect to the deductions which the Behans characterize as "impossible," the most effective refutation, it seems to me, lies in the simple fact that the deductions have been done. To carry out an experimental test of this I used a group of eight professional logicians and six

experimental psychologists. They are in agreement in performing the "impossible" deductions.

4. The Behans obviously have confused material implication with logical implication.¹

5. Their use of the formula Spec. is erroneous. This formula is not in the sentential calculus (e.g., $P \supset Q$), but in the logic of quantifiers, e.g., $x(Fx \supset Gx)$; on this, see Quine (3, pp. 17-18, ch. 2).

6. The best comments on the "proof" of the consequences of nonconfirmation for any theory are to be found in Ayer and Cohen and Nagel.²

7. The Behan paper is guilty of invalid inference. Specifically, in the final paragraph their conclusion would be formally valid if *and only if* there were a true additional premise "and everything we have said is factually true and formally correct."

¹ This confusion is evident in their equating material implication with prediction, after stating the "well-known theorem." It is not true that a false statement *logically* implies any statement, and hence it does not predict at all. The "well-known theorem" is a theorem in the calculus of propositions and it states that a false statement *materially* implies any statement. To equate this with prediction leads to absurdities which I shall demonstrate in a moment. The distinction between these two kinds of implication is given by Cohen and Nagel somewhat as follows: Logical implication "involves no assumption as to the factual truth or falsehood of either of two propositions, but only that they are so connected by virtue of their structure . . . that it is impossible for the implicating proposition to be true and the implied one to be false" (2, p. 127). It is quite different with

material implication, which is "the name we give to the fact that one of a pair of propositions happens to be false or else the other happens to be true" (2, p. 128). Of material implication Quine says: "This relation is so broad as not to deserve the name implication . . ." (3, p. 29).

The most effective way of demonstrating the consequences of this confusion is to exhibit its absurd results. Allowing the "well-known theorem" to formulate a prediction would lead to permitting a statement like the following: "Two plus two equals five predicts that Sacco and Vanzetti were executed for murder and Alfred Smith was defeated for the presidency in 1928 predicts that the base angles of an isosceles triangle are equal" (2, p. 127). Readers conversant with symbolic logic will have noticed that this is the same error which I committed in my 1951 paper. Had the Behans singled this out as a major error and confusion I should have had no choice but agreement. A second error was my failure to realize that the sentential calculus does not provide the resources for the formulation of a psychological theory (3, pp. 17-18, 65-71). This same error is also perpetuated in the Behan paper, which accepts the sentential calculus as the basis for argument.

² On this point Ayer and Cohen and Nagel are quite explicit. Ayer writes that, in the case of a nonconfirming event, "we may conclude that the [theory] is invalidated by our experiment. *But we are not obliged to adopt this conclusion* [italics mine]. If we wish to preserve our [theory] we may do so by abandoning one or more of the other relevant hypotheses" (1, p. 94). Cohen and Nagel take the same position: "The logic of the crucial experiment, therefore, is as follows: If H_1 (theory) and K (assumptions) then p_1 . But p is false; therefore *either* H_1 is false or K (in part or completely) is false" (2, p. 220). Readers who may wonder how it was so "easy" to demonstrate the reverse of this "with the help of symbolic logic" will find the answer in the previous paragraph; i.e., the Behans applied a formula from the calculus of quantifiers to a case in the sentential calculus.

REFERENCES

1. AYER, A. J. *Language, truth and logic*. London: Gollancz, Ltd., 1948.
2. COHEN, M. R., & NAGEL, E. *An introduction to logic and the scientific method*. New York: Harcourt, Brace, 1934.
3. QUINE, W. V. *Mathematical logic*. Cambridge: Harvard Univ. Press, 1951.
4. MACCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, 55, 95-107.
5. REICHENBACH, H. *Experience and prediction; An analysis of the foundations and the structure of knowledge*. Chicago: Univ. of Chicago Press, 1938.
6. WOODGER, J. H. The technique of theory construction. *Int. Encycl. unified Sci.*, 1939, 2, No. 5.

Received October 7, 1953.

SPECIAL REVIEW

AN EVALUATION OF THE ANNUAL REVIEW OF PSYCHOLOGY (VOLUMES I-IV)¹

LYLE H. LANIER
University of Illinois

One consequence of the postwar expansion of professional psychology—as science, as education, and as practice—has been an increasing volume of literature in all its branches. Beset by growing scientific specialization and professional diversification, psychologists individually have had decreasing time and competence to give to the assimilation of this profusion of publication. The important functions of review, evaluation, and systematization have lagged seriously behind the accumulation of articles, monographs, and books. Whether or not this material really represents a single universe of scientific discourse or can be transformed into a unified conceptual structure have been questions of mounting concern to psychologists.

Into this state of growing confusion the *Annual Review of Psychology* was introduced in 1950. It "was conceived as a supplement to, rather than a duplication of, other publications presenting abstracts or relatively long-term reviews of psychological literature . . . to present critical appraisals of current research and theory in psychology on an annual basis in the case of the most active and general fields and on a bi-

ennial basis for fields of lesser activity or scope." Four volumes have now appeared, edited by Calvin P. Stone with Donald W. Taylor as associate editor. The purpose of the present review is to evaluate this series as a whole, and to comment in particular upon Vol. 4.

GENERAL APPRAISAL OF THE SERIES

Although the *Annual Review* is widely known among psychologists, there are probably many who do not know in detail what the series has contained and who the contributors have been. Both for this reason and for its uses as a framework for the present review, a summary of information concerning all four volumes is presented in Table 1.

Size and bibliographic scope. The *Annual Review* has grown from 330 pages in 1950 (including indices) to 485 pages in 1953—an increase of about 47 per cent. This increase in length has roughly paralleled the number of references cited. In 1950 there were 1,594 titles in the eighteen chapters, while in 1953 the number was 2,003 for nineteen chapters. The gross increase was 38 per cent, while the average increase per chapter was about 25 per cent (from 89 to 111).

There are interesting differences among the authors in the number of references cited, both for different reviews of the same topic and for reviews of different topics. For example, Sears in 1950 listed only 39 articles on personality, while Bron-

¹ STONE, CALVIN P., & TAYLOR, DONALD W. (Eds.) *Annual review of psychology*. (4 vols.) Stanford, Calif.: Annual Reviews, Inc., 1950, 1951, 1952, 1953. Two of these volumes have been previously reviewed in this JOURNAL: Vol. 1 by H. H. Kendler in 1951 (pp. 159-161) and Vol. 3 by M. H. Marx in 1952 (pp. 657-660).

TABLE 1
SUMMARY OF CONTENTS OF *Annual Review of Psychology*

Topic	Volume 1			Volume 2			Volume 3			Volume 4		
	Author	Pp.	Refs.	Author	Pp.	Refs.	Author	Pp.	Refs.	Author	Pp.	Refs.
Devel., child	Jones & Bayley	8	38	Barker	22	69	Nowlis & Nowlis	28	136	Harris	30	160
Learning	Melton	22	101	Buxton	22	90	Harlow	26	121	Underwood	28	103
Vision	Bartlett	18	169	Chapanis	20	96	Helson	30	120	Vernon	30	154
Hearing	Newman	22	105	Wever	14	94	Garner	20	87	Licklider	22	92
Somes., chem.	Geldard	16	62	Pfaffmann	16	88	Wendt	26	172	Ruch	26	91
Ind. diff.	Thorndike	18	89	Tyler	18	91	Humphreys	20	79	Anastasi	20	139
Personality	Sears	14	39	MacKinnon	24	73	Eysenck	24	104	Brenner	26	111
Social	Bruner	32	101	Katz	36	101	Smith	30	113	Newcomb	32	117
Industrial	Shartle	22	128	Bellows	20	98	Brown & Ghiselli	28	126	Harrell	24	140
Comparative & physiological	Hebb	16	67	Deese & Morgan	24	136	Nissen & Semmes	28	157	Hess	16	71
Abnormal	Cameron	18	142	Taub	22	216	Zubin	22	68	Neff	18	107
Clinical: diagn.	H. Hunt	14	97	Challman	20	104	Magaret	38	196	White	22	75
Clinical: therapy	Snyder	14	109	Pepinsky	22	117	Raimy	30	79	Rotter	22	115
Educational	Cronbach	20	87	Stroud	24	154	Elmgren	28	104	Sanford	26	68
Counsel.: diagn.	Berdie	12	65	Stuit	12	60	Gilbert	30	120	Carter	20	129
Counsel.: therapy	Bordin	10	43	Pepinsky	18	80	McNemar	10	33	Williamson	18	99
Statist., design	Grant	20	113	Edwards	18	67				Mosteller	28	84
Problem solving	Johnson	14	39									
Gerontology				Shock	18	146						
Motivation							Mowrer	20	55			
Spec. disabil.										Cobb	26	138
Theoret. psych.										Bergmann	23	40

fenbrenner cited 111 in 1953. By contrast, Cameron mentioned 142 references in his 1950 discussion of abnormalities of behavior, while White listed 75 in 1953. Such variations are, of course, very difficult to interpret since authors differ greatly in the extent to which evaluation or even individual mention is made of numbered references.

The number of references cited in Vol. 4 is roughly 25 per cent of the number of entries in the *Psychological Abstracts* for 1952. The significance of this percentage is by no means clear, however, since the *Psychological Abstracts* cites many informational and "professional" contributions which would not properly come within the purview of the *Annual Review*. It also includes many articles in related fields which would not normally be included in an annual review of psychology. Moreover, the undetermined amount of duplication among the references in the *Annual Review* would inflate this gross index of "bibliographic coverage."

It is of some interest to compare quantity of literature reviewed in the *Psychological Bulletin* with that covered in the *Annual Review*. In 1950 some 1,363 titles were cited in articles and book reviews in the *Psychological Bulletin*, while the *Annual Review* listed 1,594 references. In 1953 the comparable figure for the *Bulletin* was 1,322 while that for the *Annual Review* had increased to 2,003. These numbers are not, of course, commensurable in any strict sense, partly because of the differing objectives of the two publications and the undetermined amount of duplication among the references in each of them.

Topical organization. It is to the chapters concerned with "basic" psychological science that one turns for the *Annual Review's* "systematic psychology"—its conceptual organization of the scientific subject matter of the discipline. Such chapters comprise approximately two-thirds of each volume. The topics fall into quite heterogeneous categories. Certain chapters are focused upon classes

of "primary" dependent variables as subject matter—general modes of behavior such as seeing, hearing, learning, and problem solving. Another set of chapters is organized in terms of "secondary" dependent variables such as indices of abilities, personality characteristics, and patterns of social behavior. A third set of chapters is concerned with different types of behaving subjects—animals, children, the aged, the disabled, the abnormal. The material on physiological psychology is partly a miscellany, but its distinctive "systematic" concern would seem to be with the influence of a set of *independent* variables upon a wide variety of behavioral functions. In this respect it is similar to those aspects of social psychology which consider the effects of social and cultural conditions upon the behavior variables serving as systematic categories for other levels of psychological analysis.

By and large, this multidimensionality of "chapter headings" is typical of the general state of contemporary systematic psychology—a state of essentially "logic-tight" compartments insofar as formal conceptual relations go. The editors of the *Annual Review* take the field pretty much as they find it, and apparently make little direct attempt to change it towards greater logical or psychological order. This judgment is not necessarily a criticism of a publication such as the *Annual Review*. It would be difficult for the editors to do otherwise than present psychology in terms of the "chapter headings," the clusters of scientists, and the classes of institutions which underlie the confused mosaic of contemporary psychology. Scientific writers probably work more easily (and willingly!) on tasks defined in terms of famil-

iar concepts and preferred categories. A related consideration is the fact that the literature tends to be organized in these terms (for example, in the *Psychological Abstracts*) and hence is more easily surveyed than would be the case if an idiosyncratic framework for systematic evaluation were established by editorial fiat. Finally, practical considerations make it virtually impossible for the editors to achieve any kind of *ex post facto* editorial coordination or integration among the several contributions. There is insufficient time between the submission of the manuscripts by the authors and the publication of the volume to accomplish much more than routine editorial emendation of the material. The various chapters cover the literature up to within six to seven months of the publication date of a given volume. To complete the writing, editing, and printing within such a period is a remarkable achievement in any case.

Nevertheless, it is unfortunate that certain focal points of contemporary psychological discussion are largely lost within the present topical structure of the *Annual Review*. Perception and motivation are examples. Volume 3 has a chapter on motivation, but none of the four volumes has had one on perception. Most of the important literature in these areas has probably been reviewed under other headings, but there have not been the continuing, unified commentaries upon the work in these fields which their importance in contemporary psychological thinking justifies. Perhaps the heterogeneous material in either of these fields could not be transformed into anything resembling a unified body of knowledge. But it is precisely because of the confusion as to the meaning and

systematic status of such terms that continuing efforts at critical methodological appraisal of their usage in research and theory would be valuable.

More extensive, independent treatment of such topics as perception, motivation, and symbolic behavior would add, of course, to the length of a volume of the *Annual Review*. If compensating reductions had to be made they might well affect the "psychophysiological" literature. Each of the four volumes has had three chapters devoted to sensory processes, whereas for perception, motivation, and thinking only the last two have appeared once each in single volumes (see Table 1). Together with special treatment of "physiological psychology," this pattern of differential emphasis overstates psychophysiological literature at the expense of strictly "psychological" material. The physiological bias is most pronounced in the chapters on somesthesia and the chemical senses. A check of the publication media represented in the bibliographies of two of these four chapters (Vols. 2 and 4) shows that approximately 80 per cent of the references are to be found in physiological and medical publications. Less frequent review of this literature would make space available for more adequate consideration of topics of greater significance to psychology as a whole.

The reviewers of Vols. 1 (Kendler) and 3 (Marx) in this JOURNAL thought that the several chapters devoted to clinical psychology and counseling should have been condensed and integrated. A single chapter for psychodiagnosis and another integrated chapter for psychotherapy has been the usual proposal for condensing these reviews. There certainly is overlap among the prob-

lems and techniques used, respectively, in "clinical psychology" and in "counseling." Indeed, this is a prominent subject of discussion by several of the reviewers, mainly those in the counseling area who are trying to find a distinctive definition of their field. But there are significant differences between these two fields in terms of literature cited in the *Annual Review*. For example, Gilbert's single chapter on counseling methods in Vol. 3 lists only 23 of the 196 references in Magaret's chapter on "clinical" psychodiagnostics and only 11 of the 79 references in Raimy's chapter on psychotherapy. The comparable figures for Vol. 4 show even less overlap: Williamson's "counseling" chapter includes only 17 of the 115 references cited in Rotter's review of "clinical" psychodiagnostics, and only two of Sanford's 68 references in the chapter on "clinical" therapy. To a considerable extent, of course, the degree of relative bibliographic independence may have been due to deliberate effort to avoid unnecessary duplication. Nevertheless, these figures, together with substantive differences between the parallel treatments, suggest that "clinical" and "counseling" psychologists at present have less over-all professional communality than the formal resemblances of their methodologies and underlying principles would seem to indicate. It might be added that in Vol. 5 the editors plan to have a single chapter on theory and techniques of assessment, one on psychotherapy and one on counseling methods. This arrangement is a step in the direction of "integration" and yet it will allow for special consideration of whatever distinctive problems and techniques there might be in the nonclinical aspects of counseling.

Evaluative character. The pattern of

topical organization provides only the structural framework for the appraisal of the literature. Other important determiners of the success of an annual review are the selection of contributors, the formulation of the reviewing task, and constructive acceptance by the reviewers of the task defined. Concerning the contributors, the list of authors in Table 1 is evidence that the editors have in general succeeded in securing psychologists from whom outstanding evaluations of the literature would be expected. The pessimists who predicted that the *Annual Review* would fail to recruit a continuing force of first-rate contributors were wrong—so far as these four volumes go.

With respect to definition of the task, the preface to Vol. 1 states: "They [the authors] were at liberty to indicate their points of departure by citing summaries of basic work prior to the review period but were advised not to strive for mere comprehensiveness, especially in the form of a loosely integrated series of abstracts of all the current literature. On the contrary, the editorial board asked that they adopt an interpretative and evaluative approach to the literature they selected for review." Thus the editors clearly foresaw the danger that the reviews might degenerate into unorganized bibliographic digests. They appeared to recognize also that certain of the reviews in Vol. 1 had failed to avoid that danger, since they remark further: "While it appears that all fields are not equally amenable to this approach, it is hoped that future volumes of the *Review* will reflect gains in the techniques of exposition and interpretation of the desired kind."

To what extent have the authors been able to achieve the general objectives set for them by the editorial

board? Have successive volumes shown improvement in the "techniques of exposition and interpretation?" Are there persistent differences among fields in the "evaluative character" of the reviews? In the attempt to answer these questions, I devised a procedure for appraising individual chapters in terms of three sets of presumably desirable characteristics: (a) an introductory orientation which defines the field and establishes a general methodological setting for the review; (b) explicit indication of the author's "systematic" approach to the field, including the rationale for his topical organization and bases of selection among references; (c) interpretative and evaluative comments upon the literature, including a summary assessment of over-all achievements and trends (e.g., in methodology, in empirical knowledge, in theory, in technology—as appropriate). In the effort to simplify the difficult judgmental task, the appraisals were restricted to "factual" estimates as to whether or not (for the first two criteria) and to what extent (for the third) the chapters exhibited these characteristics. As far as possible, differences in quality of writing, in technical sophistication, and in validity of argument were disregarded.

The results of these appraisals are represented by the figures in Table 2. In general, it appears that about one-third of the 63 chapters common to the four volumes lack the kind of "orientational" introduction described above. More than half of them omit any explicit discussion of the reviewer's "systematic" analysis or organization of the field. One-fourth have little or no interpretation or evaluation of the literature reviewed. These generalizations are subject, of course, to whatever quali-

TABLE 2
RESULTS OF RATINGS OF CHAPTERS FOR
"ORIENTATION," "SYSTEMATIZATION,"
AND "INTERPRETATION-EVALUATION"*

Classification	Volume			
	1	2	3	4
No. of chapters without "orientation"	8	6	5	3
No. of chapters without "systematic" rationale	10	12	9	9
Extent of interpretation and evaluation†				
Median rating	4	3.1	5.5	4.8
No. of chapters rated 1 or 2	5	6	2	3

* The chapters common to all four volumes were included, as well as those condensed or expanded in the last two volumes (see Table 1).

† The chapters were rated on a 7-point scale in which "1" represented the minimal and "7" the maximal degree of "interpretation-evaluation."

fications might be required by the unreliability of the method and by possible idiosyncrasies in this single reader's standards of judgment.

The editors' hope for improvement in future volumes—expressed in Vol. 1—appears to have been fulfilled, if the figures in Table 2 can be accepted as evidence. Clearly, Vols. 3 and 4 excel the first two in estimated degree of interpretation and evaluation, with Vol. 3 standing highest and Vol. 2 lowest in this respect. Two mitigating circumstances in behalf of Vols. 1 and 2 should perhaps be noted: (a) the writers of the later volumes could profit by seeing the earlier ones and from critical reactions to them on the part of the editors and other critics; (b) the last two volumes are considerably longer, thus presumably making more space available for evaluative commentary. The weight of these considerations is reduced, however, by the fact that individual reviews in Vols. 1 and 2 are rated just as high as those in Vols. 3 and 4. The incli-

nation and critical capability of the individual author are probably the principal determiners of the variance among reviews in evaluative character.

There might, nevertheless, be differences among fields in "amenability" to the kind of reviewing in question, as the editors have suggested. Although not conclusive evidence for such a question, my individual ratings show that the reviews of the applied fields are less "systematic" and "evaluative" than those in the strictly scientific areas. Again, however, the intragroup differences are much greater than the average differences between the groups. Certain of the reviews in clinical psychology stand as high by our criteria as those in any other field, in contrast to the reviews of industrial psychology. With such a broad range of application of techniques from so many branches of psychology, the literature in the latter field is quite heterogeneous, and many of the studies are highly "atomistic" in nature. Integrative evaluation and systematization are obviously easier to accomplish in an applied field such as clinical psychology where the material is organized more definitely in relation to general principles and theory.

VOLUME 4—1953

With its nineteen chapters and 485 pages (including indices), this largest member of the *Annual Review* family presents one innovation which will be welcomed by most psychologists: the inclusion in several of the bibliographies of the full titles of all references. This option was apparently available to all authors but was exercised by only seven of them. (It is disappointing to find that the chapter on special disabilities follows the

Annual Review's original practice of numbering references in the order of citation in the text.)

There are two "special" reviews: a chapter by Bergmann on theoretical psychology and one by Cobb on special disabilities. Otherwise, the organization of topics remains essentially what it was in earlier volumes (see Table 1). There are separate chapters on comparative and on physiological psychology in place of the former combined chapter, and counseling continues to have a single chapter as in Vol. 3. In general, Vol. 4 is a very impressive addition to the evaluative literature of psychology.

Child psychology. In contrast to Barker's soul-searching analysis of the uncertain and declining state of this field in Vol. 2, Harris presents a picture of buoyant empiricism, documented by the longest bibliography in the book. With so much to write about, he doesn't worry over the difficulty of defining the field rigorously—in contrast to previous reviewers. He is mainly concerned to report upon the 160 studies in the bibliography, in the form of brief summaries of method and results. There is little evaluation or systematic interpretation, and the organization of the material is poor.

Learning. Underwood's review continues the generally high level of previous reviews in this field—particularly with respect to careful interpretation and evaluation of significant studies. He excludes from consideration the host of investigations in which learning is merely involved or demonstrated, restricting the review to "work in areas with established methodology and some theoretical orientation." This means a fairly heavy concentration upon animal learning and human conditioning, an

unfortunate necessity arising from the state of the field and which Underwood probably deplores as much as anyone. Research on thinking, he notes, "continues to move at an appallingly slow pace," although "there are rumblings which portend better things to come."

Little attention is given to questions of over-all systematization of the field of learning, either by way of introductory orientation or in the main body of the review. Perhaps conceptual cross references and integration are not possible for most of these presently somewhat discrete problems, but continuing attention should perhaps be given to these desiderata on the part of reviewers in the *Annual Review* and in similar publications.

Sensory processes. Vernon's review of visual literature is the only one of the three chapters on the "senses" which presents an organizing framework for the literature to be reviewed. Licklider for hearing and Ruch for somesthetic and chemical senses get under way merely with brief indications of what their respective reviews will emphasize. Once past the introduction, however, Vernon contents herself pretty much with descriptive reporting of a wide range of literature (extending from simple threshold phenomena to the "New Look" studies of motivational factors in perception). Licklider, on the other hand, writes a highly lucid interpretative evaluation of the auditory literature. It appears to an outsider to be a balanced account of the entire spectrum of research in this field, insofar as a single year's output yields such a product. Ruch's chapter is restricted almost entirely to research on physiological mechanisms underlying somesthesia, with only about three pages on the "chemi-

cal senses." It is a well-written review, somewhat better in respect of interpretation than of evaluation.

Comparative and physiological psychology. Hess remarks about comparative psychology that it is "probably high time we stop deploring the lack of an orienting theoretical approach." He finds one in the work of European investigators, particularly Tinbergen and Lorenz. With this general orientation he proceeds to give a good account of a relatively small number of studies. Other comparative psychologists, notably Schneirla, will probably disagree with Hess's view that this recent European work is the needed "shot in the arm" for their somewhat neglected field.

Neff doesn't discuss the systematic status of physiological psychology, although he commences his review with two brief paragraphs which mention the problems of major current interest to the specialist in this field. He then proceeds to a descriptive digest of studies classified under these headings: sensory discrimination, basic drives, emotion, learning-memory-problem solving, biochemical and neuroanatomical changes in mental disease. There is almost no interpretation or evaluation.

Differential psychology and personality. Four chapters fall under these general headings: individual differences (Anastasi), personality (Bronfenbrenner), abnormalities of behavior (White), special disabilities (Cobb). All four fields are concerned with psychological differences among individuals and classes of individuals.

Anastasi stays within the psychometric framework in writing a well-organized review of the work on individual differences. The discussion is rather more on the interpretative than the evaluative side, except in

the parts concerned with the biological and social determinants of these differences. There she is critical of "hereditarian" inferences from inadequately designed research.

Bronfenbrenner's chapter is well written and it ranks high in the dimension of "interpretation-evaluation." His introductory section states clearly his "biases" and his justification for them. Recognizing that "the attempt to deal . . . with multiple, interacting relationships rather than simple static entities often leaves the theoretician with a system of ambiguous abstractions," he prefers the opportunities of the former to the narrow constraints of the latter approach.

White's chapter on abnormalities of behavior has a far more pronounced psychodynamic orientation than its predecessor by Zubin in Vol. 3. The latter stressed organic and genetic factors in his selection of references and in his general evaluative framework. White includes such material—in particular a good discussion of biochemical investigations of schizophrenia—but he emphasizes the influence of learning and life history upon abnormal behavior. He concludes with the hope that "if we can arrive at more basic biological reaction patterns and also at more basic psychological reaction patterns these two classes of events will prove at last to be truly convergent." In terms of its technical characteristics as a review, this chapter is one of the best in the book.

Cobb defines a "special disability" as any defect or disability that may occur in an otherwise normally functioning person. Her bibliography is comparatively long, but its utility for readers is greatly reduced because it is not in alphabetical order. The material reviewed is a mixture of

basic and applied research on the etiology and psychological characteristics of visual, auditory, and speech defects, together with publications on treatment and rehabilitation. The chapter is well written and is in general an excellent survey of a field not widely known to most psychologists.

Social psychology. Newcomb opens his review with an interesting innovation. By way of appraising efforts to define social psychology, he makes a comparative analysis of the content of six recent textbooks. His conclusion is that "social interaction" is the distinctive subject matter of social psychology, "that the term stands for something which can be studied at its own level." The phenomena of social interaction cannot be explained, he thinks, by "mere extrapolation of general psychological principles."

Newcomb organizes his material under two main headings: attitudes and the processes of interaction and communication—without discussing the systematic interrelationships among the problems and concepts of the two areas. The presentations of research findings are exceptionally good, although there is less critical evaluation than in Smith's chapter in Vol. 3.

Clinical psychology and counseling. Rotter sees the situation in psychodiagnosis much as Magaret did in the preceding volume: "experimental investigation of test validity emphasizes more and more the absence or inefficiency of prediction." His review is well planned, and probably has as much structural unity as a good theoretical orientation can give to an appraisal of work on a multitude of incommensurable tests. There are three broad headings: (a) general contributions to the methodology and

theory of personality measurement; (b) clinical instruments; (c) research instruments. There is a good summary at the end called "Analysis of Trends." Reviewers in other fields would do well to emulate this feature.

Sanford's account of psychotherapy is not as systematic or comprehensive as Raimy's chapter in Vol. 3, although the chapter as a whole is a balanced interpretative appraisal in which clinical and scientific values are happily blended. A good report is given of the work in group therapy at the Tavistock Clinic in London, based upon personal observations over a period of several months. The longest section in the review is devoted to research on psychotherapy, which Sanford regards as a fruitful means to the general investigation of personality.

Williamson classifies the literature on counseling into (a) publications oriented towards psychotherapeutic theory and techniques, and (b) those focused upon "the choosing of occupational goals based upon the diagnosis of aptitudes and interests." He hasn't much to say about therapeutic studies, devoting most of the review to investigations of various diagnostic instruments. There is very little interpretation or evaluation.

Educational and industrial psychology. Superficially dissimilar, these two fields have many points of methodological and technological resemblance. Both involve a broad range of psychological principles and practices designed to improve the effectiveness of fundamental social functions. Both draw upon all levels of psychological science—general, individual, and social. Perhaps for this reason, neither represents a unified scientific-technological speciality—as the two chapters in Vol. 4 will attest.

For Carter, educational psychol-

ogy deals with school learning and its correlates. His review begins with a brief outline of the problems and conditions which further define the field. Such topics as readiness for learning, acquisition of desirable attitudes, meaningful learning, sociological correlates of learning, emotional factors in learning, and measurement problems predominate. Very little work on the learning of specific school subjects per se is reported. And there is almost no relationship between this literature and that reviewed by Underwood for the general psychology of learning. On the whole, this review is a good one, in marked contrast to that of Elmgren in Vol. 3.

The lack of evaluation in the reviews of industrial psychology has already been noted. Harrell does, however, present a fairly extensive introduction in which over-all trends are effectively discussed. Major emphasis in the body of the review goes to studies of human relations in industry. As a rule these field studies appear to be inconclusive because of such factors as inadequate samples, uncontrolled conditions, and uncertain criteria.

Statistical theory and research design. Mosteller's chapter is considerably longer than its predecessors in earlier volumes, partly because it covers a wider range of topics. Special features of the review are discussions of nonparametric statistics and ranking methods, which are being used increasingly by psychologists whose data often fail to satisfy the assumptions of the more conventional parametric methods of analysis. There is an informative section in which studies dealing with the effects upon the common parametric statistics of departure from assumptions are reviewed. The general conclusion is "that departures from assumptions

may have little effect at some points of the distribution function, say near the mean or median, but far out in the tails of the distribution errors get to be large and the model may just no longer be appropriate." He suggests that it would be useful to have empirical studies of the effect of violating the assumptions to varying degrees on this model. His attitude—not unusual in mathematical statisticians of empirical bent—seems to be that while illicit inferences are to be avoided, unnecessary scrapping of information is likewise to be minimized. Questions of the utility and limitations of various other mathematical "models" are discussed in a special section, including efforts to develop probability models for learning data, uses of information theory, and difficulties in finding a suitable model for accident proneness.

Theoretical psychology. This is Bergmann's term for the logic of psychology. As a branch of the philosophy of science, theoretical psychology in this sense is concerned with the nature and structure of psychological concepts, with the laws in which they occur, and the theories into which these laws combine. Following this definition of the field, the author outlines the essential tenets of a "philosophy of psychology" on which he thinks virtual agreement has been reached, and devotes the remainder of the chapter to a critical analysis of deviations from these precepts.² First, certain dissents from this position in principle are noted: (a) the phenomenology of Snygg and Combs as regards the derivation of concepts; (b) concerning laws, London's antideterminism; (c) Skinner's attack upon theory. After brief refu-

² To designate this general methodological position, Bergmann uses the terms "neobehaviorism" or "logical behaviorism."

tations of these attacks, Bergmann gets down to the main business of the chapter which is the "prophysiological" movement in theoretical psychology. This development has two foci: (a) the Meehl-MacCorquodale-Feigl argument for the use of hypothetical constructs as distinct from intervening variables, and (b) neogestaltism as represented in recent articles by Krech. Although Bergmann's

analysis may not convince his antagonists, I think it is an outstanding contribution towards the clarification of the issues. It is perhaps unfortunate that he had to use so vulnerable a target as Krech for his attack upon gestaltism. But the gestaltists are mainly to blame; very few of them have ventured into the realm of formal methodological argument.

BOOK REVIEWS

SHAFFER, G. WILSON, & LAZARUS, RICHARD S. *Fundamental concepts in clinical psychology*. New York: McGraw-Hill, 1952. Pp. xi+540. \$6.00.

This book could well have been two separate publications. The authors have divided their efforts so that one half the chapters are written by one and the remainder by the other. There is not much similarity in approach or in style between the two sections. The first eight chapters are a sort of annotated bibliography covering the clinical field. The coverage is quite broad but not especially deep. The last seven chapters are largely devoted to psychotherapy and related topics.

The book should be useful in a number of ways, one of which will be as an aid to students preparing for comprehensive examinations. It includes a great deal of information and it directs students to other excellent sources.

It does not, however, live up to the promise of its title. Probably a representative group of clinical psychologists could not agree completely about which concepts in the field are really fundamental. But it is likely that most of them would agree that concepts such as psychological determinism, frustration-aggression, func-

tional autonomy, dynamic mechanisms, and partial reinforcement, to name a few, are central to the field. None of these is given more than cursory attention. Although psychoanalysis is discussed briefly in two different sections, the ego mechanisms are not included. The kindest thing to be said for the discussion of the Oedipus complex is that it is inadequate.

The book is written in the first person plural. The editorializing, which appears frequently, will be annoying to some readers and stimulating to others. The authors often criticize other clinicians for careless validation, slipshod methodology, and cloudy thinking. On the other hand, many readers are likely to take exception to their undocumented statements about the therapeutic process. In discussing therapy the authors stress the importance of such things as self-knowledge on the part of the therapist, the mastery of a variety of techniques, and special forms of therapeutic training. Such statements as "insight or uncovering therapy is, of course, the preferred treatment for most adjustive difficulties," or "reassurance is necessary in all therapeutic situations," or "the patient must uncover repressed material" will strike some

readers as dogmatic and unsupported by firm data.

It is likely that there would be no general agreement among clinicians that a book dealing with fundamental concepts in the field should include five chapters on psychotherapy. There would probably be fewer votes for the necessity of a chapter on physical and chemical therapies. The latter chapter is quite detailed down to citing dosage of Mitragnol and methods of applying electrodes. Much material in the earlier chapters is abbreviated or omitted on the plea of space limitation. Judgments will differ on questions of what should have been included and what left out. Many will feel that the book overemphasizes therapy at the expense of other concepts.

One final objection will occur to some readers. The authors constantly refer to persons with behavior disorders or adjustive difficulties as "patients" who "suffer" from "illnesses." This seems to imply a fundamental concept about the nature of the clinician's problem which should be further examined. Thinking about problems is frequently affected by the words used in their definition, and many psychologists feel strongly that by calling people with emotional difficulties "sick" there are created unnecessary semantic barriers to effective action.

In a work that tries to cover as much ground as this one it is easy to find faults and to disagree with the authors' choices of material for inclusion. The authors have made a real attempt to cover the field. Readers' estimates of the success of the effort will vary with their agreement with the writers' choices of topics emphasized.

GEORGE W. ALBEE.
University of Helsinki.

KLEIN, MELANIE, HEIMANN, PAULA, ISAACS, SUSAN, & RIVIERE, JOAN. (Eds.) *Developments in psychoanalysis*. London: Hogarth Press and Institute of Psycho-analysis, 1952. Pp. viii+368. 30s.

The psychoanalytic dictum that early infancy is the crucial developmental period has been a source of as much frustration as stimulation among psychologists. While no psychologist these days seriously questions the importance of this early period, neither is any psychologist particularly satisfied with the evidence for its importance or the methods for its study. The retrospective reconstruction of infancy through adult analyses, and the description of babies' feelings and attitudes by way of adult empathy, have always left psychologists dissatisfied and impatient. On the other hand, there are no conventional, generally accepted psychological methods for interpreting the emotional significance of infant behavior. Consequently, there are many theories, little data, and deep frustrations.

The present collection of papers, in adding much to theory and little to data, may strain the frustration tolerance of the psychological reader. For the most part, the book aims to clarify Melanie Klein's theoretical position concerning early infancy. As usual, her contributions are imaginative, provocative, and controversial. A fair evaluation of the writings, however, requires criticism on two different levels.

One may begin by considering the theoretical contributions alone. Here there is little question that real advance has been made in the sharpening and extension of Klein's concepts beyond the position she maintained in *The Psychoanalysis of Children* (1932). The emphasis on Freud's

concept of death instinct, and its use as the source of basic infantile anxiety, provides a starting point for the first of two phases of infant development: the *paranoid-schizoid position*. A major defense during this early period is splitting—basically the splitting between the “good” and “bad” breast, but reflecting as well a splitting or nonintegration of the ego. It is only as the ego becomes able to sustain anxiety that the infant can adopt the more mature *depressive position*, which itself fosters defenses leading to greater ego integration. The relationship of this developmental sequence (greatly oversimplified by the reviewer) to the growth of introjection and projection represents a major contribution to psychoanalytic theory.

Not everybody will agree with this expansion of theory. Many of the points made by Klein have been for years a source of controversy between one English analytic group and certain European analysts. Too often, this controversy has ended with mild name-calling: who is the “genuine” psychoanalyst and who the impostor? Frequently enough, it has involved real questions as to the appropriateness of an extension of certain Freudian concepts to the earliest months of life.

One cannot really evaluate theoretical contributions alone, however. Internal consistency is never enough. The more closely and carefully the detailed accounts of postulated infant development are written, the more they cry for empirical test. The test is not to be found in these papers, although a suggestion as to Klein's methods of arriving at hypotheses is. The instances which confirm Klein's hypotheses, as reported in one chapter, are taken from published British studies of infant development, from

her own clinical observations, and from casual, sometimes secondhand accounts of what babies do. As always, Melanie Klein in her observations is the sensitive, intuitive investigator from whom few meanings of infant behavior are hidden. Riviere and Isaacs add notably to her accounts. But there is no consideration of alternative hypotheses, no account of the negative case, no hint as to the path from observed behavior to intuitive interpretation to theory construction.

Even in the absence of these requisites of scientific method, the psychologist might be satisfied if the hypotheses advanced were testable. A careful survey of Klein's notions suggests that some probably are subject to test, although these are the peripheral rather than the crucial hypotheses. If one cannot study the content of the unconscious fantasy of a three-month-old infant, for example, one still may be able to vary the frustration to which he is subjected and observe the typical defenses. If one cannot be sure of the infant's attitude toward the breast as part-object, one may still examine the relative need-satisfying properties of food and attention. Although one may not yet be able to relate weaning to persecutory anxiety in the infant's thoughts, one may nevertheless look to the spontaneous play of weaned infants as an index of the significance of lost objects to the child.

It is an encouraging sign that numerous isolated articles in the literature of child development touch upon just such points as these. Today their number is increasing. Indeed, much of Klein's theorizing sounds less bizarre now than it did in 1932, largely because we are gradually accumulating from infants observational material which has been ordered in

psychoanalytic terms. The lasting importance of these papers may well lie not so much in the theoretical controversies they foster as in the direction they give to the systematic observation of infants.

ANN MAGARET GARNER.

*University of Illinois,
College of Medicine.*

GREENE, EDWARD B. *Measurements of human behavior.* (Rev. Ed.) New York: Odyssey Press, 1952. Pp. xxiv+790. \$4.75.

Readers familiar with the first edition of this book will be prepared for the comprehensive scope of the revision and its vast compilation of detailed and up-to-date information about measuring techniques of all varieties. They will be disappointed, however, in their expectations of a more tightly integrated book with a clearer and more precise exposition of basic principles and concepts of psychological measurement.

The fault may lie in the nature of the task the author set for himself. What was difficult in 1941 has become virtually impossible a decade later. Measurement and evaluation devices have proliferated rapidly in all areas of psychology and the development of their systematic rationale has not kept pace with empirical applications. Any book which attempts to encompass practically all the techniques (and little of the underlying theory) which have been or are employed in quantifying and assessing human behavior must suffer from some superficiality and lack of coherence. And the range covered in this one is formidable: from psychophysical methods to the Szondi test, from Thurstone's attitude scaling to OSS evaluation procedures and Kinsey's interviewing methods. Unfortunately, the book is also exasperatingly crammed

with minutiae, often irrelevant. Even the subjects who posed for the photographs of test administrations are graciously identified by name.

The introductory chapters provide only a weak foundation for supporting this mass of material. For example, the problem of the nature of measurement is dispatched in a paragraph; the procedures for developing measuring instruments are given in the form of brief exhortations such as the following: "decide specifically what is to be measured and how; . . . secure a large number of sample items . . ."; or, "analyze the responses to each item to determine such attributes as content. . . ." The other introductory chapters deal with test nomenclature (in which a curious distinction is made between achievement, aptitude, and psychological [!] tests) and the characteristics of a "good instrument." Apart from a cursory discussion of reliability, norms, and the questionable attribute of test "uniqueness," the author never develops the general methodological considerations essential to an understanding of test construction and accurate interpretation of test results. Such questions as the extent to which psychological tests satisfy the criteria for measurement, the logic underlying the concepts and empirical determination of test reliability and validity, the selection of the standardization group, etc. are disregarded or only incidentally treated. Nor are the concepts and procedures included in the three chapters on "elementary statistics" (one of which deals with factorial analysis) effectively applied to the problems involved in evaluating and using tests. In this connection, it is regrettable that the chapters in the earlier edition on "persistent problems" were deleted.

The remaining 500-odd pages are devoted primarily to a description of specific tests and related tools. There are seven chapters concerned with tests for early childhood, individual tests of ability, group tests of ability, motor and mechanical tests, tests of special aptitude, educational achievement tests, and those developed for military personnel in World War II. In each of these chapters, representative tests are described and illustrated, scoring techniques explained, and some of the research findings on the use of the test surveyed. Many new and valuable tests developed since the appearance of the first edition have been included. There are, however, a few surprising omissions like the WISC and the Merrill-Palmer whereas some comparatively ancient tests of limited value have been retained.

The chapters on attitude-, interest-, and personality-measurement have been greatly expanded. They are preceded by a distressing introduction to dynamic theory and structure of personality. The need for such a chapter is clear since many projective tests are discussed. But the oversimplified and, in places, muddled account points up the unrealistic design of the book. Clinical tests like the Bender-Gestalt, TAT, Rorschach, and Draw-a-Person are described, for example, quite fully. Such material, however, cannot serve in lieu of training manuals, nor is the methodology sufficiently explored to give the reader insight into the complex problems related to the validation as well as the clinical and research use of these tools.

Some errors of fact and naïveté are understandable in a survey text of this scope. Still, inaccuracies like the following are too frequent: "the word trait . . . is used to refer to any

physical aspect of a person . . . or to any mental aspect such as speed of reading . . ."; "the sum of the verbal test scores [on the W-B] yields a verbal MA . . ."; "this [the standard deviation] provides a method of scaling scores . . . comparable with the best physical scales . . ."; "the K score [on the MMPI] is the number of answers omitted because the client cannot say or will not choose."

The breadth of content combined with the unsystematic approach limits the audience for the book. It can probably best serve as a reference text for advanced students who wish to become acquainted with a wider variety of tests than the standard laboratory and practicum courses are able to cover. The profusion of visual illustrations, the research bibliographies, and the fairly complete classified lists of available tests and inventories should be particularly helpful for this purpose.

EVELYN RASKIN.

Brooklyn College.

BAUMGARTEN, FRANZISKA. (Ed.) *La psychotechnique dans le monde moderne.* (Psychotechnics in the modern world.) Paris: Presses Universitaires de France, 1952. Pp. xi+630. 2,000 fr.

Appropriately, these *Proceedings* of the Ninth International Congress of Applied Psychology were published in the series *Bibliothèque Scientifique Internationale* (Sciences Humaines, Section Psychologie, with H. Piéron as the chief editor), designed principally as means for acquainting the French-speaking psychologists with important developments in the field of scientific psychology. In this series were published the translations of Rorschach's *Psychodiagnostics* (together with the Böchner-Halpern volume on clinical applications of the

Rorschach test); G. H. Thomson's treatise on factorial analysis of human ability; and H. J. Eysenck's work on the dimensions of personality; and a large number of books by American authors (W. H. Sheldon's *Varieties of Human Physique*, two volumes by A. Gesell and F. Ilg, C. Wolff's *Human Hand*, as well as R. S. Woodworth's *Experimental Psychology*, C. T. Morgan's *Physiological Psychology*, and D. Krech and R. S. Crutchfield's *Social Psychology*). Among the few volumes by French authors are J. M. Faverage's book on statistical methods and a book on early child development by O. Brunet and I. Lezine.

There has been an interval of a long, eventful 15 years between the Eighth Congress of the International Psychotechnical Association (Prague, 1934) and the Ninth Congress (Bern, 1949). The editor stressed in her Introduction that the Congress has offered to the participants an opportunity to become acquainted with the present, dramatically altered features of psychotechnology.

To provide a historical perspective, Dr. Baumgarten, as the secretary general of the Association, requested reports on developments in the field of applied psychology in different countries during the war years. The communications were published under the title *Progress of Psychotechnics* (1939-1945) by A. Francke in Bern. The continued interference of political ideologies with the handiwork of the applied psychologist is reflected in the fact that the Polish delegation, in a letter of May 5, 1950, withdrew the manuscripts of papers prepared for the Congress, which they were unable to attend.

It appears as a wise policy that the *Proceedings* have not been limited to abstracts of the papers nor do they

reflect with photographic accuracy what has transpired. The papers have been frequently shortened and regrouped. Among the outstanding recent trends in psychotechnology, Dr. Baumgarten noted the concern with social problems (including human relations in industry) and the search for tests of personality.

JOSEF BROŽEK.

University of Minnesota.

LAWESHE, C. H. (Ed.) *Psychology of industrial relations*. New York: McGraw-Hill, 1953. Pp. vii + 350. \$5.50.

The stated purpose of this book is "to present in a reasonably concise and non-technical form some of the content of industrial psychology which . . . might be useful to people who work with or manage other people." The reader will find that the authors have been very successful in fulfilling this purpose. New material is presented in a style readily understandable by the supervisory force and by other management personnel. An exhaustive list of references at the end of each chapter makes it possible for the reader to investigate any topic in greater detail.

The rigorous scientific approach to problems typical of the Purdue group is evident throughout the book, and the application of the experimental method to industrial relations is documented with practical examples.

One example of the critical attitude exercised by the authors in evaluating the studies cited can be seen in the chapter on employee supervision, in which the results of a nationwide survey of foremen are reported. The original article indicated that many of the foremen felt they were not posted on company

policies and that union shop stewards and others had usurped some of their rights. The authors of this book point out that their conclusions cannot be justified by the basic data and that only a minority of foremen held these unfavorable views.

Three chapters of the book are devoted to employee and group relations. These chapters are outstanding and should prove most helpful to management personnel. The discussion of the informal group in the work force will enable foremen to understand the reasons for its existence. When the supervisors realize that the informal group insures the individual personal recognition in a highly impersonal system and supplements the formal organization structure they will be better able to carry out their duties.

Naturally, no book contains all the material the reader thinks should be included since there are practical limits of size which must be observed. Throughout this book a few chapters would have been strengthened by the addition of new material or a more thorough discussion of the material presented. The chapter on motivation and discontent in industry is somewhat oversimplified even for the nonpsychological reader.

It is implied that the industrial psychologist is able to determine the motivation for an employee's behavior by examining the work and home environment. If the reader is not aware of the effort required and the limitations of the present state of the art, he may well expect overnight changes in the work force and subsequently be disappointed when they do not occur.

In the chapter on employee counseling it is suggested that in some organizations the first-line supervisor will have to double as employee counselor. This is undoubtedly true,

but not enough attention is devoted to the fact that it would be a difficult task for the supervisor to function in this dual capacity. The supposition that the supervisor can set aside his interests and activities for the time he acts as the employee counselor is questionable.

Unfortunately, a major failing of this book is the fact that the authors ignored union-management relations under the Taft-Hartley Act. The provision for an enforced "cooling off" period under this Act has very definite implications for industrial relations. The worker, the union shop steward, and the foreman must try to "carry on" as usual during a period when negotiations would have reached a climax under other conditions. Worker-foreman relationships should also have been considered in the prestrike and poststrike periods.

This book represents an effort to fill a gap where no adequate literature previously existed. It will be very helpful to supervisory personnel and will enable many people who do not have specialized training in this area to understand and use industrial psychological techniques. It is further suggested that this book would prove valuable as a second text in management training courses.

JOSEPH W. WISSEL.

Dunlap and Associates, Inc.

POWDERMAKER, FLORENCE B., & FRANK, JEROME D. *Group psychotherapy: studies in methodology of research and therapy*. Cambridge: Harvard Univer. Press, 1953. Pp. xv+615. \$6.50.

The meat of this book lies in the numerous detailed descriptions of critical events arising in an extensive series of group therapy sessions with neurotic and schizophrenic veterans, events which have been maturely reflected upon and used as a basis for

helpful generalizations about the process and technique of group therapy. The authors and their collaborators are sensitive observers, conscientious reporters, and sensible commentators. The trimmings of the book lie in the research data reported, which are nice to have but not nearly so nourishing as the analyses of group meetings and the excellent commentaries thereon.

The report is based on a group effort lasting over a period of two years. Some 120 neurotic patients of a mental hygiene clinic and some 170 chronic schizophrenics in a psychiatric hospital were treated by 19 different psychiatrists working with a total of 27 groups. A research team of eleven persons, including psychiatrists, psychologists, and social workers, observed and recorded the meetings and then worked together to make communicable sense of the results of their experiences. An average of five patients attended the clinic groups. The groups of psychotics were more than twice as large, a circumstance reported as favoring therapy with these severely disturbed people. One group, and a fascinating one, was the entire population of a ward, about 85 patients. Most of the therapists were inexperienced in working with groups at the start of the project, and though they seem to share confidence in catharsis and interpretation as sources of gain in therapy, their newness to the group situation ensured considerable variability in approach. Groups varied in size and in rationale of composition; some groups were open, others were closed; duration of therapy varied from 8 to 175 meetings. In the neurotic group most patients received individual therapy concurrently. Such sources of variability are splendid for generating hypotheses but impose tangible limitations

on their verification. An observer sat with each group and was responsible in collaboration with the therapist for recording what went on. Tape recorders were used as an aid to accuracy but not as a source of verbatim transcripts. The accounts of sessions and of critical events were made more graphic and complete by descriptions of postures, facial expressions, and side plays, and more pointed by the omission of material considered irrelevant. But at the same time, students of therapy who have come to find verbatim transcripts indispensable will find them—indispensable.

Though weak research-wise, the book is packed, 600 pages full, with discerning observations about the dynamics of interpersonal relationships in therapy and with good, practical suggestions for making group therapy work. At this stage of our knowledge of group therapy, the problem is not so much to validate its effectiveness, which the authors essayed with limited success, as it is to define the process and to describe rigorously just what goes on. This latter need the book meets admirably, for one approach to therapy and for a particular population. It will have much immediate value for therapists and will be a rich source of hypotheses for more circumscribed and definitive studies.

NICHOLAS HOBBS.

Peabody College.

GELLHORN, ERNST. *Physiological foundations of neurology and psychiatry*. Minneapolis: Univer. of Minnesota Press, 1953. Pp. xiii + 556. \$8.50.

With the rediscovery of Freud and psychoanalysis during the thirties by organized American psychiatry, popular interest swung away from the organic etiology of mental disorder

and tended to concentrate itself upon "dynamic," developmental concepts and explanations. This break with tradition was a healthy one, but as is so often the unfortunate consequent of such extreme pendular movements within a discipline, there was not only an overemphasis upon functional factors, but a real neglect and almost a degradation of the physiological approach. It has seemed to this reviewer, however, that the last few years have seen the beginnings of a corrective movement with the pendulum swinging back toward interest in the physiological, and some hope for a more healthy balance between functional and organic influences in our study and interpretation of psychopathosis.

Professor Gellhorn's book is therefore a very timely one. Its explicit purpose is to survey and evaluate the links which can be established between experimental neurophysiology and clinical neurology and psychiatry. It begins with a survey of the basic factors regulating neuronal activity, proceeds through a discussion of the physiology and pathology of movement, the physiology of consciousness, and the physiology and pharmacology of the autonomic nervous system to a series of stimulating and provocative integrative chapters on neuro-endocrine action, the physiological basis of emotion, conditioning, homeostasis, and the phenomena of "constancy." The closing section on applications, relates the material to the clinical problems of the psychoses and psychoneuroses, and shock and carbon dioxide therapy.

This is admittedly a huge canvas, and no single author could be expected to do the picture complete justice. "Justice," however, is a relative term, and in this reviewer's opinion, Gellhorn has done an excellent job. There are some omis-

sions, and it is obvious that Gellhorn writes with greater detail and greater enthusiasm where his own experimental work is concerned. He has not hesitated to take sides on controversial issues and occasionally he has made specific suggestions for treatment. This is illustrated by the chapter on the restitution of movement after central lesions where he pummels the general concept of the "plasticity" of neural function, and offers suggested lines to be followed in re-education. The net result of his partisan attitude, however, does not interfere with the fairness or scope of his survey, and does add a flavor and a definite liveliness to his treatment. Considering the difficulty of the subject, the book is well written and reads relatively easily. Gellhorn has succeeded in his goal of demonstrating "the fruitfulness of the physiological method for the study of pathological phenomena and the rewards for physiology itself of this type of research."

For psychologists the book is timely and important. In an age which is marked by its absorption in functional explanatory concepts, a devotion to higher order statistical abstractions, and a resulting tendency to intellectualize and verbalize the problems of human behavior, Gellhorn presents fundamental physiological data whose explanatory powers and predictive potential cannot be neglected. The book holds both a promise and a warning for contemporary psychology and cannot be overlooked.

WILLIAM A. HUNT.

Northwestern University.

SZONDI, L. *Experimental diagnostics of drives.* (Translated by Gertrud Aull.) New York: Grune and Stratton, 1952. Pp. x+220. \$13.50.

For two reasons the English trans-

lation of this volume, first published in German in 1946, is something of an anticlimax. Most psychologists in this country are now generally familiar with Szondi's test and unusual theories through the work of his student and co-worker, Susan Deri. Furthermore, research results on the Szondi Test appearing in recent years are generally unfavorable. There is little need to touch on the test itself in this review despite the fact that the book contains the basic manual and, presumably, the validating evidence for the test as a psychodiagnostic instrument. Borstelmann and W. Klopfer's excellent review and critical evaluation of the research on the test appeared in the March 1953 issue of this journal. Although Susan Deri claims that the test is not dependent upon the genetic theories of its author, Szondi apparently believes that data obtained from the test results verify his theories.

In *Experimental Diagnostics of Drives* Szondi presents a highly speculative theory of personality which, he states, is derived from a union of genetics and depth psychology (psychoanalysis) validated by the results of "more than 4000 experiments." The several thousand experiments evidently refer to that many individual administrations of the Szondi Test to an unreported number of subjects from an unspecified sample of the "general population," and to undescribed criterion groups of abnormal subjects. His preferred technique for citing evidence consists in anecdotal descriptions of single cases whose test profiles always seem to coincide exactly with the principle under discussion. The text also contains a number of percentage tables, but two of these are labeled "relative frequencies (assumed)," and the others are related

in some undisclosed fashion to the "4000 experiments." Sample *N*'s, means, and dispersions are not given.

This book is definitely not concerned with research on motivation, at least as we know it. The inclusion of the word "experimental" in the title and in many chapter headings is misleading. Thus the reader is forced to assess the personality theory on the basis of its reasonableness and internal consistency. To the American psychologist, the theory's reasonableness is immediately suspect from Szondi's preface which declaims that this "new approach serves as an independent means of psychodiagnostics in the service of psychopathology, vocational psychology, psychology of delinquency, pedagogy and characterology. . . . According to the working hypothesis repressed latent genes in the lineal (inherited) unconscious determine the choice in love, friendship, profession, sickness and death (*sic*)."

The role of dominant manifest genes in this psychology of predestination is not mentioned. Motivation, he claims, is based on eight specific drive needs (factors) derived as arbitrary polarities from four independent hereditary syndromes. He implies that the syndromes are accepted by geneticists interested in mental disorder. As there is no qualitative difference in motivation between the normal and the abnormal, the drive needs must be universals.

The book *may* contain some intuitive, basic insights into human personality, but, as far as can be judged, Szondi's insights are largely Freudian constructions translated into a new and less reasonable framework.

VICTOR C. RAIMY.

University of Colorado.

BOOKS AND MONOGRAPHS RECEIVED

- DVORINE, ISRAEL. *Dvorine pseudo-isochromatic plates*. (2nd Ed.) Baltimore: Waverly Press, 1953. Pp. 28. \$12.00.
- FLOYD, W. F., & WELFORD, A. T. *Symposium on fatigue*. London: H. K. Lewis, 1953. Pp. vii+196. 24s.
- HARMS, ERNEST. *Essentials of abnormal child psychology*. New York: Julian Press, 1953. Pp. xiii+235. \$5.00.
- HOVLAND, CARL I., JANIS, IRVING L., & KELLEY, HAROLD H. *Communication and persuasion; psychological studies of opinion change*. New Haven: Yale Univer. Press, 1953. Pp. xii+315. \$4.50.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. W., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953. Pp. xxii+384. \$6.00.
- MALRIEU, PHILIPPE. *Les origines de la conscience du temps: les attitudes temporelles de l'enfant*. Paris: Presses Universitaires de France, 1953. Pp. 157.
- NICHTENHAUSER, ADOLF, COLEMAN, MARIE L., & RUHE, DAVID. *Films in psychiatry, psychology and mental health*. New York: Health Education Council, 1953. Pp. 269. \$6.00.
- NOTCUTT, BERNARD. *The psychology of personality*. New York: Philosophical Library, 1953. Pp. 259. \$4.75.
- OSBORN, ALEX F. *Applied imagination; principles and procedures of creative thinking*. New York: Scribner's, 1953. Pp. xvi+317. \$3.75.
- RHINE, JOSEPH BANKS. *New world of the mind*. New York: William Sloane, 1953. Pp. xi+339. \$3.75.
- SARTRE, JEAN-PAUL. *Existential psychoanalysis*. New York: Philosophical Library, 1953. Pp. viii+275. \$4.75. (Trans. by Hazel E. Barnes.)
- SCHEIFELE, MARIAN. *The gifted child in the regular classroom*. New York: Teachers Coll., Columbia Univer., Bureau of Publications, 1953. Pp. x+84. \$.95.
- SHAW, FRANKLIN J., & ORT, ROBERT S. *Personal adjustment in the American culture*. New York: Harper, 1953. Pp. ix+388.
- SPINELY, B. M. *The deprived and the privileged; personality development in English society*. New York: Grove Press, 1953. Pp. vii+208. \$4.00.
- STERN, ALFRED. *Sartre; his philosophy and psychoanalysis*. New York: Liberal Arts Press, 1953. Pp. xxii+223. \$4.50.
- STOLUROW, LAWRENCE M. (Ed.) *Readings in learning*. New York: Prentice-Hall, 1953. Pp. viii+555. \$6.00.
- VERNON, PHILIP E. *Personality tests and assessments*. London: Methuen, 1953. Pp. xi+220. \$4.00.
- WEITZENHOFFER, ANDRÉ M. *Hypnotism; an objective study in suggestibility*. New York: Wiley, 1953. Pp. xvi+380. \$6.00.

**THE
FOURTH
MENTAL
MEASUREMENTS
YEARBOOK**

Edited by

OSCAR K. SUMS

**Director, Institute of Mental
Measurements
Rutgers University**

- The basic source of reference.
- More than 300 contributing authorities.
- The only complete critical and analytical review of standard tests.
- Covers all commercially available tests—educational, psychological, and vocational—published or reported in English speaking countries in the period 1945 through 1951.
- Comprehensive coverage.
- Contains unbiased reviews.
- Complete, convenient indexes.

Pp. xiv, 1162. \$36.00

The Gryphon Press

Highland Park

New Jersey



Now . . . 22 outstanding scientists contribute to the new Second Edition of an internationally-known work which was described by *The American Journal of Psychiatry* as offering "unexcelled coverage of the achievements of child psychology."

John E. Anderson
Leonard Carmichael
Ruth M. Cruttschank
Karl C. Pratt
Helen Thompson
Arnold Gesell
Norman L. Munn
Florence L. Goodenough
Dorothea McCarthy
Harold E. Jones
John E. Horrocks
Margaret Mead
Vernon Jones
Arthur T. Jersild
Kurt Lewin
Sibylla Escalona
Catherine Cox Miles
Lewis M. Terman
Leona E. Tyler
Clemens E. Benda
Harold H. Anderson
Gladys L. Anderson

Based on the findings of modern research, this book demonstrates that the speculative period in child psychology is past. Completely revised, it includes three new chapters—The Adolescent, Psychopathology of Childhood, and Social Development in the Child.

MANUAL OF CHILD PSYCHOLOGY

Second Edition

Edited by **LEONARD CARMICHAEL**
Smithsonian Institution

Send for a copy on approval

1954 1295 pages \$12.00

JOHN WILEY & SONS, Inc.
440-4th Ave., New York 16

GEORGE BANTA PUBLISHING COMPANY, MENASHA, WISCONSIN